AMERICAN

JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY

AND

THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXXXII

1960

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

UNIFORMLY BOUNDED REPRESENTATIONS AND HARMONIC ANALYSIS OF THE 2×2 REAL UNIMODULAR GROUP.*[1]

By R. A. Kunze and E. M. Stein.

TABLE OF CONTENTS

---

* Received November 13, 1958.

[1] This research was supported by the United States Air Force under contract No. AF-49(638)-42, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

Introduction. This paper deals with a study of the real 2×2 unimodular group. Our study of this particular group is motivated by two factors. First, this group has an intrinsic interest, especially in view of its connection with several branches of Analysis. Secondly, the 2×2 real unimodular group affords an illuminating example for the study of other groups.

We construct a family of uniformly bounded representations of the group, and consider its implication with regard to the Fourier analysis of the group. These representations are constructed with the following properties. They all act on a *fixed* Hilbert space $\mathfrak{H}$; they are determined by a complex parameter s, $0 < R(s) < 1$, and depend analytically on the parameter s; finally, when $R(s) = \frac{1}{2}$, these representations are, up to unitary equivalence, the continuous principal series.

The above properties, in particular the analyticity, together with certain convexity arguments applied to operator valued functions yield the following:

1) The "Fourier-Laplace" transform of a function f in $L_1(G)$ exists as an operator-valued function $\mathfrak{F}$, whose values $\mathfrak{F}(s)$ act on $\mathfrak{H}$, and which is analytic in s, $0 < R(s) < 1$.

2) When $f \in L_p(G)$, $1 \leqq p < 2$, the Fourier-Laplace transform $\mathfrak{F}$ can still be defined, and is an operator valued function analytic in the strip,

$$1 - 1/p < R(s) < 1/\mathrm{p}.$$

3) A detailed analysis of the proofs of the above reveals the remarkable fact: If $f \in L_p(G)$, $1 \leqq p < 2$, the Fourier-Laplace transform $\mathfrak{F}$ of f is uniformly bounded in the operator norm along the line $R(s) = \frac{1}{2}$.

In conjunction with an analysis of the discrete series of representations, 3) implies the following significant fact concerning harmonic analysis on the group: Let k be a function in $L_p(G)$, $1 \leqq p < 2$. In contrast with the (non-compact) abelian situation, the transformation

$$f \to f * k,$$

of convolution by k, is a bounded operator on $L_2(G)$.

We shall now discuss certain of these facts in greater detail. The representations we consider arise as follows. Let

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad ad - bc = 1$$

be an element of the group. We then consider, for each complex s, the multiplier representations,[2]

[2] These representations may be put in the form originally obtained by Bargmann [1] by means of the transformation $x = \tan(\theta/2)$.

$$(1.1) \qquad f(x) \to |bx+d|^{2s-2} f((ax+c)/(bx+d))$$

$$(1.2) \qquad f(x) \to \operatorname{sgn}(bx+d)\,|bx+d|^{2s-2} f((ax+c)/(bx+d)).$$

The two continuous principal series are obtained from these by setting $s = \frac{1}{2} + it$ and restricting the functions f to lie in $L_2(-\infty, \infty)$.

We are led to the construction of the uniformly bounded representations described above by the following considerations. In the group we distinguish a particular subgroup, namely, the subgroup of lower triangular matrices of the form

$$g = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}, \quad a \neq 0.$$

It may be shown that when the representations of either of the principal series are restricted to this subgroup then they are all unitarily equivalent. This raises a natural problem. Can one find a Hilbert space $\mathfrak{H}$ and representations $U^+(\cdot, \frac{1}{2}+it)$, $U^-(\cdot, \frac{1}{2}+it)$ unitarily equivalent to (1.1), (1.2) (for $s = \frac{1}{2}+it$) such that $U^\pm(\cdot, \frac{1}{2}+it)$ when restricted to the lower triangular subgroup are independent of t? The answer is in the affirmative; furthermore, the uniformly bounded representations $U^\pm(\cdot, s)$, $0 < R(s) < 1$, which we construct, are characterized as the analytic continuations of the representations $U^\pm(\cdot, \frac{1}{2}+it)$. It should be added that the representations $U^+(\cdot, \sigma)$, $0 < \sigma < 1$, are unitarily equivalent to the complementary series.

In solving the above problem and in the actual construction of these representations, it is natural to consider the induced action of (1.1), (1.2) on the Fourier transforms F of the functions f. The considerations here are rather involved but have an intrinsic interest. For the analysis reveals connections with both the so called "Hilbert transform" and the notion of "fractional integration."

It is clear that the multiplier representations (1.1) or (1.2) afford (at least formally) an analytic continuation of (1.1) or (1.2) when $s = \frac{1}{2} + it$. The problem, however, is how to make this precise, i. e., the problem of finding an underlying Hilbert space on which these representations act and depend analytically on s. Ehrenpreis and Mautner have also dealt with the problem of extending the representations (1.1) to values of $s \neq \frac{1}{2} + it$. Their result concerning the analyticity of the Fourier transform of an L_1 spherical function on G was one of the motivating facts in our work. In addition, it was brought to our attention by Ehrenpreis that a similar result might hold for L_p. However, there are significant differences between their results and ours. In [5] they construct uniformly bounded representations arising from (1.1) when $s \neq \frac{1}{2} + it$; nevertheless, these representations act on different

Hilbert spaces depending on s. In [6], and [7] they consider, at least implicitly, representations which act on a fixed Hilbert space; but in this case the representations are not uniformly bounded when $s \neq \frac{1}{2} + it$.

The preceding considerations, in particular 3) above, lead to characterizations of the representations of the group. This may best be understood in the following context. As a result of the work of Bargmann [1], Godement [8], and Harish-Chandra [14], attention has been focused on a particular class of representations, those which are "square integrable" in the following sense; a representation $g \to U_g$ on a Hilbert space $\mathfrak{H}$ is of this type if the function

$$\phi = \phi(g) = (U_g \xi, \eta)$$

is in $L_2(G)$ for every ξ, η in $\mathfrak{H}$. The square integrable irreducible unitary representations of the 2×2 real unimodular group are essentially the representations of the discrete series. We are able to give a similar characterization of the representations of the continuous principal series. An irreducible unitary representation $g \to U_g$ is equivalent to one of the latter if and only if

$$(U_g \xi, \eta) \in L_q(G)$$

holds for all $\xi, \eta \in \mathfrak{H}$ and all $q > 2$, but not for $q = 2$. An analogous characterization holds for the representations of the complementary series. (See Theorem 10 and its corollary, in § 11.)

One of the main ideas motivating this paper was the desire to extend the classical Hausdorff-Young theorem to the group. We recall the form of the Hausdorff-Young theorem for Fourier transforms, as given by Titchmarsh. Let $f \in L_p$, $1 \leqq p \leqq 2$, and let

$$F(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ixy} f(y)\, dy.$$

Then

$$\| F \|_q \leqq (2\pi)^{\frac{1}{2}-1/p} \| f \|_p,$$

where $1/p + 1/q = 1$.

The most convenient method for proving this theorem is by using a convexity principle for linear transformations introduced by M. Riesz. This principle allows one to "interpolate" between various bounds of linear transformations. For a general discussion of this method of proof we refer the reader to [3]. An extension of the above theorem to locally compact abelian groups, via the Riesz convexity principle, is given in Weil's book [23]. An abstract generalization of this theorem to arbitrary locally compact unimodular groups has been given by one of us, [16]. This general theorem was proved

by what amounts to an extension of the convexity principle to linear transformations between operator valued functions.

Due to the analytic structure of the family of uniformly bounded representations of G, it is possible to prove a version of the Hausdorf-Young theorem which is much stronger than its classical analogue (see Theorem 7 in § 8). The proof of Theorem 7 necessitates yet another extension of the Riesz convexity principle—from the case of a single fixed linear transformation to a family of transformations depending analytically on a parameter.[3]

It seems quite likely that many of the results described above hold not only for this group but for certain other groups as well (e. g. the complex classical groups). We hope to return to this matter at a later time.

We now proceed to describe the organization of this paper.

In Chapter I, which consists of §§ 2, 3, and 4, we consider operator valued functions and we prove the basic convexity (interpolation) theorems. §§ 2 and 3 are quite general in nature. However, in § 4 the subject matter is tailored to fit the situation which arises in the 2×2 real unimodular group.

Chapter II concerns itself with the actual construction of our family of representations. In § 5 the general background and theorems are stated. Their proofs, however, require some extensive Fourier analysis. This is done in § 6. In § 7 we return to the proofs of the stated theorems.

Combining the results of Chapters I and II, we study the "Fourier-Laplace" transform for the group in Chapter III. This leads to our extension of the Hausdorff-Young theorem which is contained in § 8 . In § 9, we complete the Fourier analysis of a function on the group by a consideration of the discrete series.

Chapter IV contains some applications of the above. In § 10 we are mainly concerned with the theorem that convolution by a function in L_p, $1 \leqq p < 2$, is a bounded operator on L_2. Some implications of this result are also deduced. Finally, in § 11, we deal with characterizations of various representations of the group and with a related notion—the "extendability" of a representation to L_p.

We should like to observe that, except for some notation, the contents of Chapters I and II are independent of each other. Since Chapter I is of a more technical nature, the reader might well begin with Chapter II which deals with uniformly bounded representation of the group.

---

[3] In the case of numerical valued functions this extension was obtained by one of us in [20].

CHAPTER I. Operator Valued Functions.

2. L_p spaces of operator valued functions. In this part we prove two purely technical theorems. In these results we have ignored various possible generalizations and have restricted our attention to a rather simple situation which appears to be adequate for our purpose.

We begin by introducing enough terminology to state the theorems.

Throughout the paper $\mathfrak{H}$ will denote a complex separable Hilbert space. The ring of all bounded operators on $\mathfrak{H}$ will be denoted by $\mathfrak{B}$. If A is any non-negative operator in $\mathfrak{B}$ and $\xi_1, \xi_2, \cdots$ is any orthonormal basis of $\mathfrak{H}$, then

$$\text{(2.1)} \qquad \operatorname{tr}(A) = \sum_{n=1}^{\infty} (A\xi_n, \xi_n)$$

is non-negative and independent of the choice of basis. The bound of an A in $\mathfrak{B}$ will be denoted by $\| A \|_\infty$, and we shall put $| A | = (A^*A)^{\frac{1}{2}}$. The p-th norm of A is then given by

$$\text{(2.2)} \qquad \| A \|_p = (\operatorname{tr}(| A |^p))^{1/p},$$

where $1 \leqq p < \infty$. The letter M will always stand for a regular measure space [4] over a locally compact space with a countable basis for open sets; the underlying topological space will also be denoted by M. We shall consider functions on M whose values are bounded operators on $\mathfrak{H}$. If F is such a function, we say that F is measurable provided

$$\text{(2.3)} \qquad t \to (F(t)\xi, \eta), \qquad t \in M,$$

is a measurable numerical function on M for each pair of vectors ξ, η in $\mathfrak{H}$. If F, G are measurable, our assumptions imply the measurability of $F + G$, FG, and F^*, these being defined in the obvious way; thus for example F^* is given by $F^*(t) = [F(t)]^*$.

An operator on $\mathfrak{H}$ is said to be of finite rank if it is reduced by a finite dimensional subspace. The set of all such operators is a two sided * ideal in $\mathfrak{B}$ and will be denoted by $\mathcal{E}$.

(2.4) By a *simple function* we shall mean a function F on M to $\mathcal{E}$ having only a finite number of distinct values, each non-zero value being assumed on a set of finite measure.

[4] For a general discussion of measure theory on locally compact spaces see Halmos [10, Chapter 10].

THEOREM 1. *If F is a measurable operator valued function on M, the norms $\| F(t) \|_p$, $1 \leqq p \leqq \infty$, are measurable as functions of t, and the relations*

$$\| F \|_\infty = \operatorname*{ess\,sup}_{t \in M} \| F(t) \|_\infty \tag{2.5}$$

$$\| F \|_p = \left(\int \| F(t) \|_p{}^p \, dm(t) \right)^{1/p}, \qquad 1 \leqq p < \infty, \tag{2.6}$$

define norms relative to which the collection $L_p(M, \mathfrak{B})$ of all measurable F with $\| F \|_p < \infty$ is a complex Banach space (one identifies two functions if they differ only on null sets). Moreover, the formula

$$\int F = \int \operatorname{tr} F(t) \, dm(t) \tag{2.7}$$

is meaningful for F in $L_1(M, \mathfrak{B})$ and defines an integral satisfying

$$\left| \int F \right| \leqq \| F \|_1. \tag{2.8}$$

THEOREM 2. *If F is any measurable operator valued function on M and $1 \leqq p \leqq \infty$, there exist simple functions $S_1, S_2, \cdots$ vanishing outside compact sets such that FS_n is integrable,*

$$\lim_n \int FS_n = \| F \|_p \tag{2.9}$$

and $\| S_n \|_q \leqq 1$, $1/p + 1/q = 1$. Furthermore, if p and q are indices such that $1/r = 1/p + 1/q \leqq 1$ and F, G are measurable, then

$$\| FG \|_r \leqq \| F \|_p \| G \|_q. \tag{2.10}$$

Finally, the simple functions vanishing outside compact sets are dense in $L_p(M, \mathfrak{B})$ for all p such that $1 \leqq p < \infty$.

Except for the minor complication that we are dealing with operator valued functions, the proofs of these results involve nothing new and proceed along standard lines. We have nevertheless included most of the details for the benefit of the reader. We mention that one might obtain similar although less explicit theorems as consequences of known results from direct integral theory and the theory of non-commutative integration; however, it seems inappropriate to complicate an essentially simpler measure-theoretic situation by such considerations. Furthermore, in our application we require these results in the rather explicit and concrete form given above.

We begin by recalling some of the facts about the trace and the p-th norms mentioned earlier. As a general reference to this part, we refer to a paper [4] of Dixmier.

Let ξ, η be fixed vectors in $\mathfrak{H}$ and define $E_{\xi,\eta}$ by $E_{\xi,\eta}(\zeta) = (\zeta, \eta)\xi$, $\zeta \in \mathfrak{H}$.

An operator E is of finite rank if and only if it is a finite sum of operators of the form $E_{\xi,\eta}$ and moreover

$$\text{(2.11)} \qquad \operatorname{tr}(E_{\xi,\eta}) = (\xi,\eta).$$

Let $\mathfrak{B}_p$, $1 \leqq p \leqq \infty$ denote the collection of all bounded operators A on $\mathfrak{H}$ such that $\| A \|_p < \infty$. A positive operator A is in $\mathfrak{B}_p$, $1 \leqq p < \infty$, if and only if its spectrum $\lambda_1, \lambda_2, \cdots$ is discrete and

$$\sum_{n=1}^{\infty} \lambda_n{}^p < \infty.$$

In this event, $\| A \|_p = (\sum_{n=1}^{\infty} \lambda_n{}^p)^{1/p}$, which implies,

(2.12) *$\| A \|_p$ is a non-decreasing function of p.*

(2.13) *If $A \in \mathfrak{B}_p$, $B \in \mathfrak{B}_q$, and $1/r = 1/p + 1/q \leqq 1$, then $\| AB \|_r \leqq \| A \|_p \| B \|_q$, where $1 \leqq p, q \leqq \infty$.*

(2.14) *If $A \in \mathfrak{B}$ and $1 \leqq p \leqq \infty$, there exist operators $E_1, E_2, \cdots$ of finite rank such that $\| E_n \|_q \leqq 1$ and*

$$\lim_n \operatorname{tr}(AE_n) = \| A \|_p,$$

where $1/p + 1/q = 1$. In case A is of finite rank, there exists another such operator E with $\| E \|_q \leqq 1$ and $\operatorname{tr}(AE) = \| A \|_p$.

(2.15) *$\mathfrak{B}_p$ is a Banach space under the norm given by (2.2), and the collection $\mathcal{E}$ of operators of finite rank is dense in $\mathfrak{B}_p$ for $1 \leqq p < \infty$.*

(2.16) *If $1 \leqq p < \infty$ and $B \in \mathfrak{B}_q$, where $1/p + 1/q = 1$, then*

$$A \to \operatorname{tr}(AB), \qquad A \in \mathfrak{B}_p$$

is a bounded linear functional, ϕ_B on $\mathfrak{B}_p$, and the map $B \to \phi_B$ identifies $\mathfrak{B}_q$ with the conjugate space of $\mathfrak{B}_p$.

Given an orthonormal basis $\xi_1, \xi_2, \cdots$ of $\mathfrak{H}$ we form the set $\mathfrak{D}$ of all finite rational linear combinations of the operators E_{ij} and $(-1)^{\frac{1}{2}}E_{ij}$, where

$$\text{(2.17)} \qquad E_{ij} = E_{\xi_i,\xi_j}.$$

$\mathfrak{D}$ is denumerable, and one easily verifies that the product of two members of $\mathfrak{D}$ is again in $\mathfrak{D}$.

LEMMA 1. *$\mathfrak{D}$ is dense in $\mathfrak{B}_p$, $1 \leqq p < \infty$.*

Suppose E is an operator of finite rank and that P is a projection of finite rank such that $EP = E$. Then for $1 \leqq p < \infty$ and A, B in $\mathfrak{D}$ we have

$$\| E - AB \|_p \leqq \| EP - AB \|_1 \leqq \| E - A \|_2 \| B \|_2 + \| A \|_2 \| P - B \|_2.$$

The inequalities follow from (2.12) and (2.13). Now because $\mathfrak{D}$ is dense in B_2, we see that any operator of finite rank can be approximated in the p-th norm by elements of $\mathfrak{D}$. An application of (2.15) finishes the proof.

As a corollary we obtain the fact that $\mathfrak{B}_p$, $1 \leqq p < \infty$, is separable.

The collection of all measurable operator valued functions on M will be denoted by $\mathfrak{B}(M)$.

LEMMA 2. *If $F \in \mathfrak{B}(M)$ and E is an operator of finite rank,*

$$t \to \operatorname{tr}(F(t)E)$$

is a measurable function on M.

There exist vectors $\xi_1, \eta_1, \cdots, \xi_n, \eta_n$ in $\mathfrak{H}$ such that

$$E = \sum_{n=1}^{n} E_{\xi_i, \eta_i}.$$

Thus $F(t)E = \sum_{n=1}^{n} E_{F(t)\xi_i, \eta_i}$ and by (2.11),

$$\operatorname{tr}(F(t)E) = \sum_{n=1}^{n} (F(t)\xi_i, \eta_i),$$

which implies $\operatorname{tr}(F(t)E)$ is measurable as a function of t.

LEMMA 3. *If $1 \leqq p \leqq \infty$ and $F \in \mathfrak{B}(M)$, $t \to \| F(t) \|_p$ is a measurable function on M.*

Let A belong to $\mathfrak{B}$. From (2.13), (2.14), and Lemma 1 we see that

$$(2.18) \qquad \| A \|_p = \operatorname{Sup}\{ | \operatorname{tr}(AE) | : E \in \mathfrak{D}, \| E \|_q \leqq 1 \},$$

where $1/p + 1/q = 1$. Replacing A by $F(t)$ and applying Lemma 2 we see that $\| F(t) \|_p$ is measurable; this follows from the fact that the least upper bound of a countable collection of measurable functions is again measurable.

LEMMA 4. *A function F on M to $\mathfrak{B}_p$, $1 \leqq p < \infty$, is measurable if and only if $t \to \operatorname{tr}(F(t)B)$ is measurable on M for all B in $\mathfrak{B}_q$ $(1/p + 1/q = 1)$.*

For every pair of vectors ξ, η in $\mathfrak{H}$, $B = E_{\xi,\eta}$ is in $\mathfrak{B}_q$; hence $F \in \mathfrak{B}(M)$, provided $\operatorname{tr}(F(t)B)$ is measurable as a function of t for all B in $\mathfrak{B}_q$.

Conversely, suppose $F \in \mathfrak{B}(M)$. Let $B \in \mathfrak{B}_q$. By (2.13) $tr(F(t)B)$ exists and is finite for each t in M. If $p = 1$,

$$\operatorname{tr}(F(t)B) = \lim_{n\to\infty} \sum_{i=1}^{n} (F(t)B_{\xi_i,\xi_i}),$$

where $\xi_1, \xi_2, \cdots$, is any orthonormal basis of $\mathfrak{H}$, and is therefore measurable. If $1 < p < \infty$, there exist, by (2.15), operators $E_1, E_2, \cdots$, of finite rank such that $\| B - E_n \|_q \to 0$. Thus

$$| \operatorname{tr}(F(t)B) - \operatorname{tr}(F(t)E_n) | \leqq \| F(t) \|_p \| B - N_n \|_q \to 0.$$

By Lemma 2, $\operatorname{tr}(F(t)E_n)$ is measurable in t for each n, and hence $\operatorname{tr}(F(t)B)$ is also.

The result just established together with (2.16) shows that a measurable function F on M to $\mathfrak{B}_p$ is weakly measurable as a function on M to the separable Banach space $\mathfrak{B}_p$; thus F is also strongly measurable.[5]

The proof of Theorem 1 now follows from the preceding lemmas and the well known theory[5] of the Lebesgue integral extended to functions with values in a Banach space.

In proving Theorem 2 it is convenient to establish

LEMMA 5. *Suppose* $F = \sum_{i=1}^{n} f_i A_i$, *where* f_i *is a measurable numerical function on* M, $f_i f_j = 0$, $i \neq j$, *and* $A_i \in \mathfrak{B}$. *Then for* $1 \leqq p < \infty$,

$$\| F \|_p = (\sum_{i=1}^{n} \| f_i \|_p{}^p \| A_i \|_p{}^p)^{1/p}, \tag{2.19}$$

and in case $\| F \|_1 < \infty$,

$$\int F = \sum_{i=1}^{n} (\int f_i(t)\, dm(t)\, (\operatorname{tr} A_i). \tag{2.20}$$

Since $f_i(t) f_j(t) = 0$ $(i \neq j)$,

$$\begin{aligned} \| F \|_p{}^p &= \int \| F(t) \|_p{}^p \, dm(t) \\ &= \int (\sum_{i=1}^{n} | f_i(t)^p \| A_i \|_p{}^p) \, dm(t) \\ &= \sum_{i=1}^{n} (\int | f_i(t) |^p \, dm(t)) \| A_i \|_p{}^p. \end{aligned}$$

If $\| F \|_1 < \infty$, then

$$\begin{aligned} \int F &= \int \operatorname{tr}(F(t))\, dm(t) \\ &= \int \operatorname{tr}(\sum_{i=1}^{n} f_i(t) A_i)\, dm(t) \end{aligned}$$

[5] For a discussion of these points see Hille and Phillips [15, Chapter 3].

$$= \int \Big(\sum_{i=1}^{n} f_i(t) \operatorname{tr} A_i \Big) dm(t)$$

$$= \sum_{i=1}^{n} \Big(\int f_i(t)\, dm(t) \Big) \operatorname{tr}(A_i).$$

LEMMA 6. *The simple functions vanishing outside compact sets are dense in $L_p(M, \mathfrak{B})$ for $1 \leqq p < \infty$.*

Let f, g be measurable numerical valued functions on M, and let $A, B \in \mathfrak{B}$. By simple estimates and the preceding lemma we have

$$\| fA - gB \|_p \leqq \| f - g \|_p \| A \|_p + \| g \|_p \| A - B \|_p, \tag{2.21}$$

where $1 \leqq p < \infty$. Thus if $\epsilon > 0$, $A \in \mathfrak{B}_p$, and if f is the characteristic function of a set of finite measure, we can choose a compact set with characteristic function g and an operator B of finite rank such that

$$\| fA - gB \|_p < \epsilon. \tag{2.22}$$

The conclusion of the lemma follows from the fact[5] that finite linear combinations of functions of the form fA are dense in $L_p(M, \mathfrak{B})$.

LEMMA 7. *If F is a simple function with compact support there exists a simple function S with compact support such that $\| S \|_q \leqq 1$, and*

$$\int FS = \| F \|_p, \tag{2.23}$$

where $1 \leqq p \leqq \infty$, $1/p + 1/q = 1$.

There exist operators $A_1, \cdots, A_n$ of finite rank and mutually disjoint measurable subsets with characteristic functions $f_1, \cdots, f_n$ such that

$$F = \sum_{i=1}^{n} f_i A_i.$$

By (2.14) there exists an operator E_i of finite rank such that $\| E_i \|_q \leqq 1$ and $\operatorname{tr}(A_i E_i) = \| A_i \|_p$. Let

$$c_i = \| F \|_p^{1-p} \| A_i \|_p^{p-1} \tag{2.24}$$

and put

$$S = \sum_{i=1}^{n} c_i f_i E_i. \tag{2.25}$$

Then

$$\int FS = \int \sum_{i=1}^{n} c_i f_i (A_i E_i)$$

$$= \sum_{i=1}^{n} c_i \Big(\int f_i(t)\, dm(t) \operatorname{tr}(A_i E_i) = \sum_{i=1}^{n} c_i \| f_i \|_p^p \| A_i \|_p = \| F \|_p.$$

Also

$$\| S \|_q{}^q = \sum_{i=1}^{n} c_i{}^q \| f_i \|_q{}^q \| E_i \|_q{}^q$$

$$\leqq \sum_{i=1}^{n} \| F \|_p{}^{(1-p)q} \| A_i \|_p{}^{(p-1)q} \| f_i \|_p{}^p$$

$$= \| F \|_p{}^{-p} \sum_{i=1}^{n} \| f_i \|_p{}^p \| A_i \|_p{}^p = 1.$$

Finally, since F has compact support, so does S.

Proof of Theorem 2. Suppose F, G are measurable. To establish (2.10) we use (2.13) which implies

$$\| FG \|_r \leqq \int (\| F(t) \|_p \| G(t) \|_q)^r \, dm(t).$$
$$\leqq \int (\| F(t) \|_p{}^p \, dm(t))^{1/p} (\int \| G(t) \|_q{}^q \, dm(t))^{1/q}$$

provided $p \neq \infty$ and $q \neq \infty$. The other two cases arse treated by similar arguments.

As the case $p = \infty$ is somewhat exceptional and requires separate treatment, we shall prove (2.9) only for p such that $1 \leqq p < \infty$.[6] Suppose first of all that $\| F \|_p < \infty$. By Lemma 6, there exist simple functions $F_1, F_2, \cdots$, with compact supports such that $\| F - F_n \|_p \to 0$. Choose S_n for F_n in accordance with Lemma 7. By (2.10) FS_n is integrable and

$$| \int FS_n - \int F_n S_n | \leqq \| F - F_n \|_p \to 0.$$

If $\| F \|_p = \infty$, let F_n be the product of F and the characteristic function of a set of finite measure contained in $\{t \colon \| F(t) \|_p \leqq n\}$ and chosen so that $\| F_n \|_p \to \infty$. Then $\| F_n \|_p < \infty$. Thus we can choose a simple function S_n with compact support contained in the support of F_n such that, $\| S_n \|_q \leqq 1$ and

$$| \int F_n S_n - \| F \|_p | < 1/n.$$

Then $FS_n = F_n S_n$ and

$$\int FS_n \to \| F \|_p.$$

3. Interpolation in the general case. In this section we prove a rather general interpolation theorem for operator valued functions. Let Φ be a complex valued function whose domain contains a strip, $\alpha \leqq Rz \leqq \beta$. We shall say that Φ is *admissible* on the strip if Φ is analytic in $\alpha < Rz < \beta$, continuous in $\alpha \leqq Rz \leqq \beta$, and satisfies the growth condition

[6] We do not need the exceptional case $p = \infty$ in our application.

(3.1) $$\operatorname*{Sup}_{\alpha\leqq x\leqq\beta} \log|\Phi(x+iy)| \leqq Ce^{\mu|y|},$$

where C and μ are constants depending on Φ; we require also that μ satisfies the additional condition

(3.2) $$\mu<\pi/(\beta-\alpha).$$

If Φ_1, Φ_2 are admissible on a given strip and if ν_1, ν_2 are complex numbers it is easily verified that the combinations $\nu_1\Phi_1+\nu_2\Phi_2$, $\Phi_1\Phi_2$ are also admissible.

A complex valued function on a measure space will be called a *simple* function if it can be expressed as a finite linear combination of characteristic functions of measurable sets of finite measure.

Now let M_1, M_2 be measure spaces, and let D be a strip, $\alpha\leqq Rz\leqq\beta$. Suppose B_z, $z\in D$, is a complex valued bilinear form defined for all simple functions f_1, f_2 on M_1, M_2. We shall say that the collection $\{B_z\}$ is an *admissible* family of bilinear forms on D if

(3.3) $$\Phi(z)=B_z(f_1,f_2)$$

is admissible on D for each pair of simple functions, f_1, f_2 on M_1, M_2.

We now introduce some notation and terminology which will remain fixed throughout this part.

The strip $\alpha\leqq Rz\leqq\beta$ will be denoted by D and we shall put

(3.4) $$\gamma=(1-\tau)\alpha+\tau\beta, \qquad 0<\tau<1.$$

We suppose p_0, p_1, q_0, q_1 are given indices such that $1\leqq p_i, q_i\leqq\infty$, and $q_0\neq\infty$ or $q_1\neq\infty$. The indices p, q are then determined by

(3.5) $$1/p=(1-\tau)1/p_0+\tau 1/p_1,$$

(3.6) $$1/q=(1-\tau)1/q_0+\tau 1/q_1.$$

The conjugate indices of q_0, q_1, q will be denoted by q_0', q_1', q'. Finally, A_0, A_1 will denote non-negative functions such that

(3.8) $$\log A_i(y)\leqq Ae^{\delta|y|}, \qquad \delta<\pi/(\beta-\alpha).$$

With minor changes, the proof of Theorem 1 [20] yields the following convexity principle.

LEMMA 8. *Let $\{B_z\}$ be an admissible family of bilinear forms on D, and suppose*

(3.9) $$|B_{\alpha+iy}(f_1,f_2)|\leqq A_0(y)\,\|f_1\|_{p_0}\,\|f_2\|_{q_0'}$$

$$(3.10) \qquad |B_{\beta+iy}(f_1,f_2)| \leqq A_1(y)\,\|f_1\|_{p_0}\,\|f_2\|_{q_1'}$$

for all simple functions f_1, f_2 on M_1, M_2. Then for simple functions f_1, f_2 we also have

$$(3.11) \qquad |\beta_\gamma(f_1,f_2)| \leqq A_\tau\,\|f_1\|_p\,\|f_2\|_{q'}.$$

The constant A_τ is given explicitly, in terms of the Poisson kernel for the strip, by

$$(3.12) \qquad \begin{aligned} \log A_\tau = \int_{-\infty}^{\infty} &\log A_0[(\beta-\alpha)y]\,w(1-\tau,y)\,dy \\ &+ \int_{-\infty}^{\infty} \log A_1[(\beta-\alpha)y]\,w(\tau,y)\,dy, \end{aligned}$$

where

$$w(\tau,y) = \tfrac{1}{2}\sin\pi/(\cosh\pi y + \cos\pi).$$

By a bounded subset of a regular measure space we mean any measurable subset of a compact set. Now let N be an arbitrary measure space. Suppose T_z, $z \in D$, is a linear transformation from simple functions f on N to measurable operator valued functions $T_z(f) = F_z$ on M. We shall say that $\{T_z\}$ is an *admissible* family on D if $(F_z(t)\xi,\eta)$ is locally integrable on M and

$$(3.13) \qquad \Phi(z) = \int_K (F_z(t)\xi,\eta)\,dm(t)$$

is admissible on D for each choice of vectors ξ, η in $\mathfrak{H}$, simple function f on N, and bounded subset K of M.

THEOREM 3. *Let N be a measure space, and suppose $\{T_z\}$, $z \in D$, is an admissible family of linear transformation from simple functions f on N to measurable operator valued functions $T_z(f) = F_z$ on M. Suppose further that the following two conditions are satisfied for each simple function f.*

$$(3.14) \qquad \|T_{\alpha+iy}(f)\|_{q_0} \leqq A_0(y)\,\|f\|_{p_0}.$$

$$(3.15) \qquad \|T_{\beta+iy}(f)\|_{q_1} \leqq A_1(y)\,\|f\|_{p_1}.$$

Then it is also true that

$$(3.16) \qquad \|T_\gamma(f)\|_q \leqq A_\tau\,\|f\|_p.$$

In proving the theorem it is convenient to establish the following lemma.

LEMMA 9. *If $\{T_z\}$, $z \in D$, is an admissible family and S is a simple operator valued function on M which vanishes outside a compact set in M. Then $\operatorname{tr} F_z(t)S(t)$ is integrable and*

$$\Phi(z) = \int \operatorname{tr} F_z(t) S(t)\, dm(t) \tag{3.17}$$

is admissible on D for each simple f on N.

Suppose first that $S = kE_{\xi,\eta}$, where k is the characteristic function of a bounded set. Then

$$\operatorname{tr} F_z(t) S(t) = k(t)\,(F_z(t)\xi, \eta),$$

and the result follows by assumption. The general case follows by linearity.

Proof of the theorem. Our assumptions imply $q \neq \infty$.[7] Thus to show that

$$\| T_\gamma(f) \|_q \leqq A_\tau \| f \|_p$$

it suffices, in view of Theorem 2, to show that

$$\left| \int \operatorname{tr} F_\gamma(t) S(t)\, dm(t) \right| \leqq A_\tau \| f \|_p \| S \|_{q'} \tag{3.18}$$

for each simple function S vanishing outside a compact set.

The idea of the proof is to reduce this problem to one concerning an admissible family of bilinear forms. We shall then apply Lemma 8 to complete the argument.

Suppose then that S is a simple function with compact support. We can express S as $\sum_{i=1}^{n} k_i E_i$, where $k_1, k_2, \cdots, k_n$ are the characteristic functions of mutually disjoint bounded subsets K_i and each E_i is an operator of finite rank. Now let $E_i = U_i \,|\, E_i \,|$ be the canonical polar decomposition of E_i, and let

$$\sum_j \lambda_{ij} P_{ij}, \qquad \lambda_{ij} > 0, \tag{3.19}$$

be the spectral decomposition of $|\, E_i \,|$. The pairs of indices i, j will then range over a finite set which we shall call M_2. To each complex valued function $g = \{g_{ij}\}$ defined on M_2 we associate an operator valued function G on M which is given by

$$G(t) = \sum_{i=1}^{n} k_i(t) \sum_j g_{ij} U_i P_{ij}. \tag{3.20}$$

Then G is a simple function with compact support, and by an elementary computation we get

$$G^*(t)\, G(t) = \sum_{i,j} k_i(t) \,|\, g_{ij} \,|^2 P_{ij}. \tag{3.21}$$

---

[7] The case $q = \infty$ could be dealt with by a more involved argument.

Now for $1 \leqq p < \infty$, $\| G \|_p^p = \int \mathrm{tr}[(G^*(t)G(t))^{p/2}]dm(t)$ which implies

$$\| G \|_p = (\sum_{i,j} | g_{ij} |^p \| k_i \|_p^p)^{1/p}. \tag{3.22}$$

Since $\| k_i \|_p^p$ is independent of p being, in fact, equal to the measure of K_i, we can introduce a measure in M_2 relative to which $\| g \|_p = \| G \|_p$. We observe that this relation is also valid for $p = \infty$. Because the maps $f \to F_z$, $g \to G$ are linear it follows that the equation

$$B_z(f, g) = \int \mathrm{tr}(F_z(t) G(t)) \, dm(t) \tag{3.23}$$

defines a bilinear form for each z in D. In this formula f is an arbitrary simple function on $M_1 = N$, and g is any complex valued, obviously simple, function on M_2. By Lemma 9, in particular by (3.17), $\{B_z\}$ is an admissible family on D. Now

$$| B_{\alpha+iy}(f, g) | \leqq \| F_{\alpha+iy} G \|_1 \leqq \| F_{\alpha+iy} \|_{q_0} \| G \|_{q_0'},$$

and using (3.14) we get

$$| B_{\alpha+iy}(f, g) | \leqq A_0(y) \| f \|_{p_0} \| g \|_{q_0'}. \tag{3.24}$$

By similar estimates we obtain

$$| B_{\beta+iy}(f, g) | \leqq A_1(y) \| f \|_{p_1} \| g \|_{q_1'}. \tag{3.25}$$

Thus by Lemma 8, $| B_\gamma(f, g) | \leqq A_\tau \| f \|_p \| g \|_{q'}$. As this holds for all g, we may take $g = \{\lambda_{ij}\}$. Then $G = S$ and

$$B_\gamma(f, g) = \int \mathrm{tr}\, F_\gamma(t) S(t) \, dm(t), \tag{3.26}$$

which implies (3.18).

4. The main interpolation theorem. In order to prove our results for the 2×2 real unimodular group, we use, in addition to facts about the group and its representations, certain convexity arguments. The basic and most important fact along these lines is established in this section; with the intent of clarifying the situation, we have presented it in a slightly more general form than our application requires.

An operator valued function $\mathfrak{F}$ defined on an open strip, $\alpha_0 < Rs < \beta_0$, in the complex s-plane is said to be *analytic* if $(\mathfrak{F}(s)\xi, \eta)$ is analytic for all ξ, η in $\mathfrak{H}$. We shall say that $\mathfrak{F}$ is of *admissible growth* in the strip if

$$\sup_{\alpha_0 < \sigma < \beta_0} \log \| \mathfrak{F}(\sigma + it) \|_\infty \leqq C e^{\mu |t|}, \qquad \mu < \pi/(\beta_0 - \alpha_0). \tag{4.1}$$

THEOREM 4. *Let N be a measure space and T be a linear map from*

simple functions on N *to analytic operator valued functions such that* $\mathfrak{F} = \dot{T}f$ *is of admissible growth on the strip* $\alpha_0 < Rs < \beta_0$ *for each simple function* f. *Suppose that for* $\alpha_0 < \alpha < \beta < \beta_0$ *we have*

$$\sup_{-\infty<t<\infty} \| \mathfrak{F}(\alpha + it) \|_\infty (1 + |t|)^c \leqq A_0 \| f \|_1, \tag{4.2}$$

$$\left(\int_{-\infty}^{\infty} \| \mathfrak{F}(\beta + it) \|_2^2 \, |t|^{2a} (1 + |t|)^{2b} \, dt \right)^{\frac{1}{2}} \leqq A_1 \| f \|_2 \tag{4.3}$$

for all simple f, *where* a, b, c *are real and* $a \geqq 0$. *Then we may conclude*

$$\left(\int_{-\infty}^{\infty} \| \mathfrak{F}(\gamma + it) \|_q^q (1 + |t|)^{qd} \, dt \right)^{1/q} \leqq A_\tau \| f \|_p, \tag{4.4}$$

where $1 < p < 2$, $1/p + 1/q = 1$, $\gamma = \alpha + \tau(\beta - \alpha)$, $d = c + \tau(a + b - c)$, *and the parameter* τ *is determined by* $1/p = 1 - \tau/2$.

Remarks. Before we prove the theorem, we notice that the result (4.4) is intermediate—in the sense of Riesz-Thorin convexity—between the hypotheses (4.2) and (4.3). It should be noted that the singularity at $t = 0$ of the measure $|t|^{2a}(1 + |t|)^{2b} dt$ does not persist in the conclusion; only the influence of $|t|^{2a}$ for t near infinity remains.

The proof given below could be generalized in several directions. We may begin with a general pair of indices (p_0, q_0), (p_1, q_1) instead of $(1, \infty)$ and $(2, 2)$. We might also consider more general measures than those of the form $|t|^{2a}(1 + |t|)^{2b} dt$ given above. We shall not consider these generalizations here.

It should be pointed out that the proof given below would be much simpler if a, b, and c were zero. In that case the left-hand sides of (4.2) and (4.3) would be translation invariant in t. Since the basic method of the proof consists of translation along vertical lines of the strip, we are forced to overcome the lack of translation invariance by somewhat complicated devices.

At several points in the proof it will be convenient to refer to the easily verified result given below:

LEMMA 10. *If* ν *is real and* $\delta > 0$, *there exists a constant* $A > 0$ *such that*

$$(\delta + |y + t|)^\nu \leqq A(1 + |y|)^{|\nu|}(1 + |t|)^\nu \tag{4.5}$$

for $-\infty < y, t < \infty$.

Proof of the theorem. We shall obtain the result as a consequence of Theorem 3. To do this we set $M = (-\infty, \infty)$ and put

$$dm = (1+|t|)^{2(a+b-c)}\,dt, \tag{4.5}$$

where dt is Lebesgue measure. Given a simple function f on N we form $\mathfrak{F} = Tf$ and set

$$F_z(t) = \mathfrak{F}(z+it)(1+|t|)^{c-a}(z-\beta+it)^a \tag{4.6}$$

for $\alpha \leqq Rz \leqq \beta$. Since $a \geqq 0$, we may choose a single valued branch of the factor $(z-\beta+it)^a$ which is analytic in $\alpha < Rz < \beta$ and continuous on $\alpha \leqq Rz \leqq \beta$. Thus $(F_z(t)\xi, \eta)$ is analytic in z for each t and is jointly continuous in z, t for all vectors ξ, η in $\mathfrak{H}$. Furthermore, the transformation T_z defined by $T_z(f) = F_z$ is linear and maps simple functions on N to measurable operator valued functions on M.

We shall now estimate $\| F_z(t) \|_\infty$. By (4.6) and the condition (4.1) that $\mathfrak{F}$ is of admissible growth in $\alpha \leqq Rz \leqq \beta$, we find,

$$\begin{aligned} \| F_z(t) \|_\infty &= \| \mathfrak{F}(x+i(y+t)) \|_\infty (1+|t|)^{c-a} |x-\beta+i(y+t)|^a \\ &\leqq A \| \mathfrak{F}(x+i(y+t)) \|_\infty (1+|t|)^c (1+|y|)^a \end{aligned} \tag{4.7}$$

Hence,

$$\log \| F_z(t) \|_\infty \leqq Ce^{\mu|y+t|} + \log(1+|t|)^c + \log(1+|y|)^a + \log A.$$

This estimate together with the above implies the condition (3.13) that $\{T_z\}$ be an admissible family. Now, for $z = \alpha + iy$ we find, using (4.2), (4.7), that

$$\begin{aligned} \| F_{\alpha+iy}(t) \|_\infty &\leqq A \| \mathfrak{F}(\alpha+i(y+t)) \|_\infty (1+|t|)^c (1+|y|)^a \\ &\leqq A \| f \|_1 (1+|y+t|)^{-c} (1+|t|)^c (1+|y|)^a. \end{aligned}$$

Thus by Lemma 10, we obtain (4.8).

$$\| T_{\alpha+iy}(f) \|_\infty \leqq A(1+|y|)^{|a-c|} \| f \|_1. \tag{4.8}$$

Next we shall estimate $\| T_{\beta+iy}(f) \|_2$. We have,

$$\| F_{\beta+iy} \|_2^2 = \int_{-\infty}^{\infty} \| \mathfrak{F}(\beta+i(y+t)) \|_2^2 |y+t|^{2a} (1+|t|)^{2c-2a}\, dm.$$

Now making use of (4.3), we obtain

$$\begin{aligned} &\| F_{\beta+iy} \|_2^2 \\ &\leqq (A_1 \| f \|_2)^2 \sup_{-\infty<t<\infty} [(1+|y+t|)^{-2b}(1+|t|)^{2c-2a}(1+|t|)^{2(a+b-c)}] \\ &\leqq (A_1 \| f \|_2)^2 \sup_{-\infty<t<\infty} [(1+|y+t|)^{-2b}(1+|t|)^{2b}]. \end{aligned}$$

Thus by Lemma 10,

$$\| F_{\beta+iy} \|_2 \leqq A(1+|y|)^{|b|} \| f \|_2. \tag{4.9}$$

Having (4.8) and (4.9) we can apply Theorem 3 and conclude that

$$\| T_\gamma(f) \|_q \leqq A_\tau \| f \|_p. \tag{4.10}$$

Now

$$\| T_\gamma(f) \|_q{}^q = \int_{-\infty}^{\infty} \| \mathfrak{F}(\gamma+it) \|_q{}^q (1+|t|)^{qc-qa} |\gamma-\beta+it|^{qa}\, dm$$
$$\leqq A_\tau \| f \|_p{}^q.$$

Hence

$$\int_{-\infty}^{\infty} \| \mathfrak{F}(\gamma+it) \|_q{}^q (1+|t|)^{qc-qa} (1+|t|)^{qa} (1+|t|)^{2(a+b-c)}\, dt \leqq A \| f \|_p{}^q.$$

Since $\tau q = 2$,

$$qd = gc + q\tau(a+b-c) = qc + 2(a+b-c).$$

Thus

$$\int_{-\infty}^{\infty} \| \mathfrak{F}(\gamma+it) \|_q{}^q (1+|t|)^{qd}\, dt \leqq A \| f \|_p{}^q,$$

which proves the theorem.

5. Uniformly bounded representations. We now consider the group G of 2×2 real unimodular matrices, and we first recall some of the known facts concerning the representations of G.

We represent an element $g \in G$, by

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad ad - bc = 1,$$

and denote by $g(x)$ the fractional linear transformation

$$g(x) = (ax+c)/(bx+d), \qquad -\infty < x < \infty.$$

Then

$$(g_1 g_2)(x) = g_2(g_1 x) \tag{5.1}$$

and $dg(x)/dx = (bx+d)^{-2}$, $bx+d \neq 0$.

We now introduce two "multipliers" ϕ^+ and ϕ^-. These are defined by

$$\phi^+(g,x,s) = |bx+d|^{2s-2} \tag{5.2}$$

$$\phi^-(g,x,s) = \operatorname{sgn}(bx+d)\phi^+(g,x,s), \tag{5.3}$$

where s is an arbitrary complex number.

Next we consider the "multiplier representations"

$$g \to v^{\pm}(g, s)$$

given for functions f on the real axis by

$$v^{\pm}(g, s) : f(x) \to \phi^{\pm}(g, x, s) f(g(x)). \tag{5.4}$$

From these, one may obtain the irreducible unitary representations of G. They fall into three classes.[8]

a) The two *continuous principal series*

$$g \to v^{\pm}(g, 1/2 + it), \qquad -\infty < t < \infty,$$

where the Hilbert space is the space L_2 of square integrable functions on $-\infty < x < \infty$, with the usual measure.

b) The *complementary series*

$$g \to v^{+}(g, \sigma), \qquad 0 < \sigma < 1/2.$$

The Hilbert space, in this case, is defined by the inner product

$$(f, h)_\sigma = a_\sigma \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) h(y) \, | x - y |^{-2\sigma} \, dx dy, \tag{5.5}$$

where $a_\sigma = \Gamma(2\sigma) \cos(\sigma\pi)/\pi$.

c) The two *discrete series*,

$$g \to D^{\pm}(g, k), \qquad k = 0, 1, 2, \cdots.$$

We shall not need the exact form of these representations.

The Plancherel formula for G was derived by Harish-Chandra [13]. It involves representations of type a) and c) and not of type b). To state it we make the usual definition

$$U(f) = \int_G f(g) U(g) \, dg$$

for uniformly bounded representations $g \to U(g)$ and f in $L_1(G)$. Using this notation the Plancherel formula asserts that, whenever $f \in L_1(G) \cap L_2(G)$,

$$\| f \|_2^2 = 1/2 \int_{-\infty}^{\infty} \| v^{+}(f, 1/2 + it) \|_2^2 \, t \tanh \pi t \, dt$$
$$+ 1/2 \int_{-\infty}^{\infty} \| v^{-}(f, 1/2 + it) \|_2^2 \, t \coth \pi t \, dt$$

---

[8] Except for notation, these representations are those of Bargmann [1]; the difference of notation is discussed more fully in the proof of Theorem 10 in § 11.

$$+\sum_{k=0}^{\infty} \| D^{+}(f,k) \|_2^2 (k+1/2)$$

$$+\sum_{k=0}^{\infty} \| D^{-}(f,k) \|_2^2 (k+1).$$

Here $\| \cdot \|_2$ means the usual Hilbert-Schmidt norm for operators as used in § 2 above.

One of our main results is contained in the following theorem.

THEOREM 5. *There exists a separable Hilbert space* $\mathfrak{H}$ *and representations*

$$g \to U^{\pm}(g,s)$$

of G *on* $\mathfrak{H}$ *with the following properties:*

1) $g \to U^{\pm}(g,s)$ *is a continuous representation of* G *on* $\mathfrak{H}$ *for each complex* s *in the strip* $0 < R(s) < 1$.

2) $g \to U^{\pm}(g, 1/2 + it)$ *is unitarily equivalent to the representation* $g \to v^{\pm}(g, 1/2 + it)$ *of the continuous principal series defined above.*

3) $g \to U^{+}(g,\sigma)$, $0 < \sigma < 1/2$ *is unitarily equivalent to the representation* $g \to v^{+}(g,\sigma)$ *of the complementary series.*

4) *If* ξ *and* η *are two vectors in* $\mathfrak{H}$, *then the functions*

$$s \to (U^{\pm}(g,s)\xi, \eta), \quad g \text{ fixed},$$

are analytic in $0 < Rs < 1$.

5) $$\sup_{g} \| U^{\pm}(g,s) \|_{\infty} \leqq A_{\sigma}(1 + |t|)^{\frac{1}{2}},$$

$s = \sigma + it$, $0 < \sigma < 1$.[9] *Furthermore, the constant* A_σ *is bounded on any interval of the form* $0 < \alpha \leqq \sigma \leqq \beta < 1$.

It is known that for each t, the representations $v^{\pm}(\cdot, 1/2 + it)$ and $v^{\pm}(\cdot, 1/2 - it)$ are unitarily equivalent. Hence the same fact holds for the representations $U^{\pm}(\cdot, 1/2 + it)$ and $U^{\pm}(\cdot, 1/2 - it)$.

As the next theorem shows, these equivalences are to some extent already inherent in the "analytic structure" of the representations $g \to U^{\pm}(g,s)$; the theorem also describes some adidtional, and rather interesting, relations among the representations $U^{\pm}(\cdot, s)$.

THEOREM 6. *The following symmetries exist:*

---

[9] As in § 2, the ordinary bound of an operator A is denoted by $\| A \|_{\infty}$.

1) *The representations* $U^+(\cdot, s)$ *and* $U^+(\cdot, 1-\bar{s})$ *are contragredient. Similarly,* $U^-(\cdot, s)$ *and* $U^-(\cdot, 1-\bar{s})$ *are contragredient.*

2) $U^+(\cdot, s) = U^+(\cdot, 1-s)$. *Hence* $U^+(\cdot, s)$ *and* $U^+(\cdot, \bar{s})$ *are also contragredient.*

3) *There exists a fixed non-scalar unitary operator* S *such that for all* s *in* $0 < Rs < 1$,

$$SU^-(\cdot, s)S^{-1} = U^-(\cdot, 1-s).$$

Thus $U(\cdot, \bar{s})$ *is unitarily equivalent to the contragredient of* $U^-(\cdot, s)$.

Remarks. (i) It should be observed that the known result [1] concerning the reducibility of the representation $U^-(\cdot, 1/2)$ is implied by 3).

(ii) The representations $U^+(\cdot, s)$ for $s \neq 1/2 + it$ and $s \neq \sigma$ are unitarily equivalent to representations introduced by Mautner and Ehrenpreis [5]. These they show are not equivalent to unitary ones. They also assert that the representations are uniformly bounded. However, the more definite statement contained in 5) of Theorem 5 is crucial for our purposes.

(iii) The proof of Theorem 5 is lengthy and requires some vigorous classical Fourier analysis. This is contained in § 6, which is, for the most part, somewhat technical. At first reading the reader may prefer to pass on to § 7.

6. Some lemmas from Fourier analysis. We shall begin by introducing a class of Hilbert spaces, which will be seen [10] to be related to the L_p spaces via the Fourier transform. These spaces $\mathfrak{H}_\sigma$, are indexed by a parameter $0 < \sigma < 1$, and are given by the norm

$$(6.1) \qquad \| F \|_\sigma = \left(\int_{-\infty}^{\infty} | F(x) |^2 \, | x |^{2\sigma - 1} \, dx \right)^{\frac{1}{2}}.$$

The spaces $\mathfrak{H}_{\sigma_1}$, $\mathfrak{H}_{\sigma_2}$ corresponding to any pair of indices σ_1, σ_2, such that $0 < \sigma_1, \sigma_2 < 1$, are naturally related by a family of unitaries which we shall now exhibit. Let $s_1 = \sigma_1 + it_1$ and $s_2 = \sigma_2 + it_2$ where $-\infty < t_1, t_2 < \infty$. Now let $W(s_1, s_2)$ be the mapping with the domain $\mathfrak{H}_{\sigma_1}$ given by

$$(6.2) \qquad F(x) \to F(x) \, | x |^{s_1 - s_2}, \qquad F \in \mathfrak{H}_{\sigma_1}.$$

[10] Although many of the results of this section are probably known, they do not seem to be accessible in the literature in the manner in which we need them.

Then

$$\| W(s_1, s_2)F \|_{\sigma_2}{}^2 = \int_{-\infty}^{\infty} | F(x) |^2 \, | x |^{2(\sigma_1 - \sigma_2)} \, | x |^{2\sigma_2 - 1} \, dx = \| F \|_{\sigma_1}{}^2.$$

This fact together with (6.2) shows that $W(s_2, s_1)$ is the inverse of $W(s_1, s_2)$.

In what follows, we shall be mainly concerned with the pair of spaces $\mathfrak{H}_\sigma$ and $\mathfrak{H}_{1-\sigma}$. For the sake of convenience we shall set $W_s = W(s, 1-s)$. The mapping W_σ is of particular interest because it implements a duality between $\mathfrak{H}_\sigma$ and $\mathfrak{H}_{1-\sigma}$. In order to make this statement precise, we shall introduce some additional notation. Throughout this section and the one that follows, it will frequently be convenient to put

$$(F, G) = \int_{-\infty}^{\infty} F(x) \overline{G(x)} \, dx. \tag{6.3}$$

This notation will be used with the understanding that F, G are measurable complex valued functions defined on $-\infty < x < \infty$ such that $F\bar{G}$ is integrable. The inner product in $\mathfrak{H}_\sigma$ will be denoted by $(\cdot, \cdot)_\sigma$, and we shall sometimes set $1 - \sigma = \sigma'$.

LEMMA 11. *If* $F \in \mathfrak{H}_\sigma$ *and* $G \in \mathfrak{H}_{\sigma'}$, $0 < \sigma < 1$, then

$$(W_\sigma F, G)_{\sigma'} = (F, G) = (F, W_{\sigma'} G)_\sigma, \tag{6.4}$$

$$| (F, G) | \leqq \| F \|_\sigma \, \| G \|_{\sigma'}. \tag{6.5}$$

Furthermore, if A *is a bounded operator on* $\mathfrak{H}_\sigma$, *the operator*

$$A' = W_\sigma A^* W_\sigma^{-1}, \tag{6.6}$$

where A^* *is the (Hilbert space) adjoint of* A *is characterized as the unique bounded operator on* $\mathfrak{H}_{\sigma'}$ *such that*

$$(A(F), G) = (F, A'(G)) \tag{6.7}$$

for all F *in* $\mathfrak{H}_\sigma$ *and all* G *in* $\mathfrak{H}_{\sigma'}$.

To prove (6.4) we first observe that $2\sigma' - 1 = 1 - 2\sigma$. Thus

$$\begin{aligned}
(W_\sigma F, G)_{\sigma'} &= \int_{-\infty}^{\infty} F(x) | x |^{2\sigma - 1} \overline{G(x)} \, | x |^{1-2\sigma} \, dx \\
&= (F, G) \\
&= \int_{-\infty}^{\infty} F(x) \overline{G(x) \, | x |^{1-2\sigma}} \, | x |^{2\sigma - 1} \, dx \\
&= (F, W_{\sigma'} G)_\sigma.
\end{aligned}$$

Now, by Schwartz's inequality,

$$|(F,G)| \leqq \| W_\sigma F \|_{\sigma'} \| G \|_{\sigma'},$$

and (6.5) follows from the fact that W_σ is an isometry. Suppose A is a bounded operator on $\mathfrak{H}_\sigma$, and that $F \in \mathfrak{H}_\sigma$, $G \in \mathfrak{H}_{\sigma'}$. By (6.4), the fact that W_σ preserves inner products, and a second application of (6.4) we find that

$$\begin{aligned}(A(F),G) = (A(F), W_{\sigma'}G)_\sigma & \\ &= (F, A^*W_{\sigma'}G)_\sigma = (W_\sigma F, W_\sigma A^* W_{\sigma'}G)_{\sigma'} \\ &= (F, W_\sigma A^* W_{\sigma'}G).\end{aligned}$$

Thus (6.7) is satisfied by the operator $A' = W_\sigma A^* W_\sigma^{-1}$. That A' is the unique operator with this property follows from the easily established fact that, $G \in \mathfrak{H}_{\sigma'}$ and $(F,G) = 0$ for all $F \in \mathfrak{H}_\sigma$ implies $G(x) = 0$ a.e. It should also be observed that $(A')' = A$.

In addition to the $\mathfrak{H}_\sigma$ spaces we shall consider the L_p spaces, $1 \leqq p < \infty$, of functions f defined on $-\infty < x < \infty$ and normed by

$$\| f \|_p = \left(\int_{-\infty}^{\infty} | f(x) |^p \, dx \right)^{1/p}.$$

Since the parameter σ ranges between 0 and 1 and $1 \leqq p < \infty$ there should be no confusion between the norms $\| \cdot \|_\sigma$, and $\| \cdot \|_p$.

For a function f defined on $-\infty < x < \infty$, the Fourier transform F is defined by

$$F(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ixy} f(y) \, dy \tag{6.6}$$

and the inverse Fourier transform is given by

$$f(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{ixy} F(y) \, dy. \tag{6.7}$$

Here and throughout this section we shall adhere to the following convention. Pairs of functions which are related to each other by either (6.6) or (6.7) will be denoted by corresponding lowercase and capital letters such as f, F or g, G. Furthermore, we take for granted such standard facts as the Plancherel theorem, and the sense in which these transforms exist for functions in L_p, $1 \leqq p \leqq 2$, as well as the equivalence of (6.6) with (6.7) under suitable restrictions on f or F. (See e.g. [21]). To be more specific, we shall make use of two well known results on L_p transforms, a theorem of Titchmarsh (the so-called Hausdorff-Young theorem), and the Parseval formula for L_p, L_q. These results may be stated as follows:

LEMMA 12. *If $f, G \in L_p$, $1 \leqq p \leqq 2$, and their Fourier transforms F, g are given by* (6.6), (6.7), *then*

a) $$\| F \|_q \leqq A \| f \|_p,$$

b) $$\| g \|_q \leqq A \| G \|_p,$$

where $1/p + 1/q = 1$, *and*

c) $$(f, g) = (F, G).$$

Now the relation between the L_p spaces and the $\mathfrak{H}_\sigma$ spaces mentioned earlier is contained in the following lemma.

LEMMA 13. *Let $f \in L_p$, $1 < p \leqq 2$, and let F denote its Fourier transform* (6.6). *Let $\sigma = 1 - 1/p$. Then $0 < \sigma \leqq 1/2$, $F \in \mathfrak{H}_\sigma$, and*

$$\| F \|_\sigma \leqq A_\sigma \| f \|_p, \qquad 1 < p \leqq 2. \tag{6.8}$$

The class of $F \in \mathfrak{H}_\sigma$ which are Fourier transforms of $f \in L_p$ is dense in $\mathfrak{H}_\sigma$; $1 < p \leqq 2$, $0 < \sigma \leqq 1/2$.

Analogously, let $F \in \mathfrak{H}_\sigma$, $1/2 \leqq \sigma < 1$, and let $\sigma = 1 - 1/p$. Then $2 \leqq p < \infty$, the inverse transform (6.7) *exists in L_p norm, and*

$$\| f \|_p \leqq A_\sigma \| F \|_\sigma, \qquad 1/2 \leqq \sigma < 1.$$

Consider first the case $1 < p \leqq 2$. By a theorem of Hardy and Littlewood (see [11], p. 375),

$$\left(\int_{-\infty}^{\infty} | F(x) |^p \, | x |^{p-2} \, dx \right)^{1/p} \leqq A_p \| f \|_p, \qquad 1 < p \leqq 2.$$

Now,

$$\| F \|_\sigma^2 = \int_{-\infty}^{\infty} | F(x) | \cdot | F(x) | \, | x |^{2\sigma - 1} \, dx$$

$$\leqq \left(\int_{-\infty}^{\infty} | F(x) |^q \, dx \right)^{1/q} \left(\int_{-\infty}^{\infty} (| F(x) | \, | x |^{2\sigma - 1})^p \, dx \right)^{1/p},$$

by Holder's inequality. Furthermore, $(2\sigma - 1)p = p - 2$. Thus using the inequalities of Titchmarsh (Lemma 12) and Hardy and Littlewood we obtain

$$\| F \|_\sigma^2 \leqq A_p \| f \|_p^2, \qquad 1 < p \leqq 2.$$

This proves (6.8).

If F is the characteristic function of a finite interval, then f given by (6.7) is in L_p for all $p > 1$. Hence finite linear combinations of characteristic functions of finite intervals are contained among the Fourier transforms (6.6) of $f \in L_p$. Therefore the image of L_p, $1 < p \leqq 2$, under the Fourier

transform is dense in the corresponding space $\mathfrak{H}_\sigma$, $\sigma = 1 - 1/p$. This concludes the consideration of the case $1 < p \leqq 2$.

The second part of the lemma, which deals with the case $1/2 \leqq \sigma < 1$, $2 \leqq p < \infty$, follows from the first part by duality. We shall briefly indicate the argument. Put $\sigma' = 1 - \sigma$. Then $\sigma' = 1 - 1/p'$, where $1/p + 1/p' = 1$ and $1 < p' \leqq 2$. By Lemma 11, $\mathfrak{H}_{\sigma'}$ and $\mathfrak{H}_\sigma$ are dual and it is well known that $L_{p'}$ and L_p are dual. The second part of the lemma then follows upon identifying (6.7) with the adjoint of (6.6) considered as a mapping from $L_{p'}$ to $\mathfrak{H}_{\sigma'}$ (properly speaking, this can be done only on a dense subset of $\mathfrak{H}_\sigma$).

LEMMA 14. *Let*

$$K(F)(x) = \int_{-\infty}^{\infty} |1 - |x/y|^{\alpha}|\, |x-y|^{-1} F(y)\, dy.$$

Then

$$\int_{-\infty}^{\infty} |K(F)(x)|^2\, dx \leqq A_\alpha \int_{-\infty}^{\infty} |F(x)|^2\, dx$$

$$\text{if } -1/2 < \alpha < 1/2.$$

This lemma is known. The proof follows easily from Theorem 319 of [12]. There, a more general theorem on integral operators whose kernels are homogeneous of degree -1 is given.

The main discussion of this section is contained in the lemma below. We shall deal with operators acting on $\mathfrak{H}_\sigma$. It will be convenient, however, to specify the action of these operators by exhibiting their action on the Fourier transforms of the functions in question.

Thus we consider the multiplication operators

$$m_t^+ : f(x) \to |x|^{2it} f(x) \tag{6.9}$$

$$m_t^- : f(x) \to \operatorname{sgn}(x) |x|^{2it} f(x). \tag{6.10}$$

Now if F is the Fourier transform of $f \in L_1 \cap L_2$, we shall denote the Fourier transforms of $m_t^+(f)$, $m_t^-(f)$ by $M_t^+(F)$, $M_t^-(F)$.

It will also be convenient to introduce the following class $\mathfrak{D}$ of functions: $F \in \mathfrak{D}$ if F is C^∞ and vanishes in a neighborhood of zero and outside a compact set. Clearly $\mathfrak{D}$ is dense in each $\mathfrak{H}_\sigma$, $0 < \sigma < 1$. Furthermore, $\mathfrak{D}$ is contained in the image of $L_1 \cap L_2$ under the Fourier transform, (6.6).

LEMMA 15. *If $F \in \mathfrak{D}$, then $M_t^\pm(F) \in \mathfrak{H}_\sigma$ for each σ such that $0 < \sigma < 1$. and the transformations*

$$F \to M_t^\pm(F), \qquad F \in \mathfrak{D}$$

have unique bounded extensions to all of $\mathfrak{H}_\sigma$.

The extensions, which will also be denoted by M_t^+, M_t^- are unitary on $\mathfrak{H}_{\frac{1}{2}}$ and, in general, the bound, $\| M_t^{\pm} \|_\sigma$ of $M_t^{\pm}$ considered as an operator on $\mathfrak{H}_\sigma$ satisfies

$$\| M_t^{\pm} \|_\sigma \leqq A_\sigma (1 + | t |)^{\frac{1}{2}}, \qquad 0 < \sigma < 1. \tag{6.11}$$

Since the restrictions of m_t^+, m_t^- to L_p, $1 \leqq p < \infty$, map L_p isometrically onto L_p, the Plancherel theorem implies that $M_t^{\pm}$ extend to unitary operators on $\mathfrak{H}_{\frac{1}{2}}$.

To treat the case $\sigma \neq 1/2$, one would like to express M_t^+ and M_t^- as integral operators of convolution type. However, the kernels in question are not locally integrable, and we must therefore proceed rather indirectly.[11]

We introduce the transformation

$$m_t^\epsilon \colon f(x) \to | x |^{-\epsilon + 2it} f(x), \qquad 0 < \epsilon < 1/2,$$

which maps $L_1 \cap L_2$ into $L_1 \cap L_2$.

Putting F for the Fourier transform (6.6) of $f \in L_1 \cap L_2$, we denote the Fourier transform of $m_t^\epsilon(f)$ by $M_t^\epsilon(F)$. Now let $F \in \mathfrak{D}$. Since F is the Fourier transform of an $f \in L_1 \cap L_2$, we can form $M_t^\epsilon(F)$, and, as is easily verified, by the Plancherel theorem,

$$\| M_t^\epsilon(F) - M_t^+(F) \|_2 \to 0, \text{ as } \epsilon \to 0. \tag{6.12}$$

Next, we claim that

$$M_t^\epsilon(F) = a_{\epsilon,t} \int_{-\infty}^{\infty} F(y) \, | x - y |^{\epsilon - 1 - 2it} \, dy \tag{6.13}$$

for F in $\mathfrak{D}$, where

$$a_{\epsilon,t} = \tfrac{1}{\pi} \Gamma(1 - \epsilon + 2it) \cos[\pi/2 (1 - \epsilon + 2it)].$$

This follows from the fact [2, p. 43] that the Fourier transform of

$$e^{-b|x|} \, | x |^{-\epsilon + 2it}, \qquad b > 0$$

is

$$(2\pi)^{-\frac{1}{2}} \Gamma(1 - \epsilon + 2it) [(b + ix)^{\epsilon - 1 - 2it} + (b - ix)^{\epsilon - 1 - 2it}]. \tag{6.14}$$

This converges to

$$(2\pi)^{-\frac{1}{2}} a_{\epsilon,t} \, | x |^{\epsilon - 1 - 2it} \tag{6.15}$$

[11] The following observations may help clarify the situation. When $t = 0$, M_t^+ reduces to the identity transform. This may be regarded as convolution by the Dirac kernel. When $t = 0$, M_t^- reduces to the so-called "Hilbert transform," which apart from a constant factor may be viewed as a convolution by the function $1/x$. In this case our result was proved by Hardy and Littlewood [11], whose argument we extend to the general case.

as $b \to 0^+$, and is bounded by $A \mid x \mid^{\epsilon-1}$, with A independent of b. Now (6.13) follows by standard convergence theorems and the Plancherel theorem.

Together with $F \in \mathfrak{D}$, consider $G(x) = \mid x \mid^{\sigma-\frac{1}{2}} F(x)$. Since $F \in C^\infty$ and vanishes in a neighborhood of zero and outside of a compact set, the same may be said of $G(x)$. Thus we may apply formula (6.10) to G as well. Call

$$\Delta_\epsilon(x) = M_t^\epsilon(G)(x) - \mid x \mid^{\sigma-\frac{1}{2}} M_t^\epsilon(F)(x). \tag{6.16}$$

Then by (6.13),

$$\Delta_\epsilon(x) = a_{\epsilon,t} \int_{-\infty}^{+\infty} [\mid y \mid^{\sigma-\frac{1}{2}} F(y) - \mid x \mid^{\sigma-\frac{1}{2}} F(y)] \mid x - y \mid^{\epsilon-1-2it} dy. \tag{6.17}$$

It is easy to verfy, (by the Lebesgue dominated convergence theorem) that

$$\| \Delta_\epsilon(x) - \Delta_0(x) \|_2 \to 0, \text{ as } \epsilon \to 0, \ (F \in \mathfrak{D}). \tag{6.18}$$

If we use (6.12) with G in place of F, (6.18) and (6.16), we obtain

$$\| \mid x \mid^{\sigma-\frac{1}{2}} M_t^+(F) \|_2 \leqq \| M_t^+(G) \|_2 + \| \Delta_0(x) \|_2. \tag{6.19}$$

As has already been noted

$$\| M_t^+(G) \|_2 = \| G \|_2$$

while

$$\| G \|_2 = \| \mid x \mid^{\sigma-\frac{1}{2}} F \|_2 = \| F \|_\sigma,$$

and

$$\| \mid x \mid^{\sigma-\frac{1}{2}} M_t^+(F) \|_2 = \| M_t^+(F) \|_\sigma.$$

Substituting the above in (6.19) leads to

$$\| M_t^+(F) \|_\sigma \leqq \| F \|_\sigma + \| \Delta_0(x) \|_2. \tag{6.20}$$

It remains therefore to estimate $\| \Delta_0(x) \|_2$. Recalling (6.17) we have

$$\Delta_0(x) = a_{0,t} \int_{-\infty}^{+\infty} \mid y \mid^{\sigma-\frac{1}{2}} F(y) - \mid x \mid^{\sigma-\frac{1}{2}} F(y) \mid x - y \mid^{-1-2it} dy$$

$$= a_{0,t} \int_{-\infty}^{+\infty} 1 - (\mid x \mid / \mid y \mid)^{\sigma-\frac{1}{2}} \mid x - y \mid^{-1-2it} \mid y \mid^{\sigma-\frac{1}{2}} F(y) \, dy.$$

We now apply Lemma 14, with $\alpha = \sigma - \frac{1}{2}$, and with $\mid y \mid^{\sigma-\frac{1}{2}} F(y)$ in place of $F(y)$. We then have

$$\| \Delta_0(x) \|_2 \leqq A_\sigma \mid a_{0,t} \mid \| \mid x \mid^{\sigma-\frac{1}{2}} F(x) \|_2 = A_\sigma \mid a_{0,t} \mid \| F(x) \|_\sigma.$$

However

$$a_{0,t} = (1/\pi) \Gamma(1 + 2it) \cos \tfrac{1}{2}\pi(1 + 2it).$$

Hence by well-known estimates in the theory of the Γ function, see [22], p. 151, it follows that

$$|a_{0,t}| \leqq A(1+|t|)^{\frac{1}{2}}.$$

Combining this with the above we obtain:

$$\|\Delta_0(x)\|_2 \leqq A_\sigma(1+|t|)^{\frac{1}{2}}\|F\|_\sigma.$$

Together with (6.20), this implies

$$\|M_t^+(F)\|_\sigma \leqq A_\sigma(1+|t|)^{\frac{1}{2}}\|F\|_\sigma.$$

This was our desired result for M_t^+.

The proof for M_t^- is very similar. The only change that occurs is that we use the fact that the Fourier transform of $\operatorname{sgn}(x)|x|^{-\epsilon+2it}$ is

$$(2\pi)^{\frac{1}{2}}b_{\epsilon,t}\operatorname{sgn}(x)|x|^{\epsilon-1-2it},$$

where

$$b_{\epsilon,t} = (i/\pi)\Gamma(1-\epsilon+2it)\sin\tfrac{1}{2}\pi[1-\epsilon+2it].$$

This concludes the proof of the lemma.

LEMMA 16. *The estimates for M_t^+ and M_t^- may be strengthened as follows. Let $\epsilon > 0$, then*

$$\|M_t^+\|_\sigma \leqq A_{\sigma,\epsilon}(1+|t|)^{(1+\epsilon)|\sigma-\frac{1}{2}|}, \qquad 0<\sigma<1,$$

with $A_{\sigma,\epsilon}$ independent of t.

Proof. Let us consider M_t^+, and assume that $\frac{1}{2} \leqq \sigma < 1$; the other cases are treated analogously. We have already noted that M_t^+ is unitary on $\mathfrak{H}_{\frac{1}{2}}$. Thus we have

$$(6.21) \qquad \left(\int_{-\infty}^{+\infty}|M_t^+(F)|^2\,dx\right)^{\frac{1}{2}} = \left(\int_{-\infty}^{+\infty}|F|^2\,dx\right)^{\frac{1}{2}}.$$

By the lemma we have just proved, we have, if $\frac{1}{2} \leqq \sigma_0 < 1$,

$$(6.22) \qquad \begin{aligned} &\left(\int_{-\infty}^{+\infty}|M_t^+(F)|^2|x|^{2\sigma_0-1}\,dx\right)^{\frac{1}{2}} \\ &\qquad \leqq A_{\sigma_0}(1+|t|)^{\frac{1}{2}}\left(\int_{-\infty}^{+\infty}|F(x)|^2|x|^{2\sigma_0-1}\,dx\right)^{\frac{1}{2}}. \end{aligned}$$

Notice that the above inequalities are of the same nature, except for the weight functions which determine the measures in question.

Now it is possible to "interpolate" between these two inequalities, and obtain intermediate ones from them. Of course we have already used many

variants of this type of argument in § 3 and § 4 above. The particular theorem we need is contained in [20], (Theorem 2). To apply it we argue as follows:

Choose σ_0, so that $\sigma < \sigma_0 < 1$. We may then write

$$2\sigma - 1 = (1-\theta)\cdot 0 + \theta(2\sigma_0 - 1) = \theta(2\sigma_0 - 1),$$

with $0 < \theta < 1$. Notice that in the above, $\sigma = \frac{1}{2}$ when $\theta = 0$, and $\sigma = \sigma_0$ when $\theta = 1$. The result of applying Theorem 2 of [20] is

$$\left(\int_{-\infty}^{+\infty} | M_t^+(F) |^2 \, | x |^{2\sigma-1} \, dx\right)^{\frac{1}{2}} \leqq A_{\sigma_0}{}^{\theta}(1 + | t |)^{\theta/2}\left(\int_{-\infty}^{+\infty} | F(x) |^2 \, | x |^{2\sigma-1} \, dx\right)^{\frac{1}{2}}.$$

However,

$$\theta = (2\sigma - 1)/(2\sigma_0 - 1).$$

Thus we choose σ_0 close enough to 1 so that $\theta \leqq (2\sigma - 1)(1 + \epsilon)$. Hence the result becomes

$$\left(\int_{-\infty}^{+\infty} | M_t^+(F) |^2 \, | x |^{2\sigma-1} \, dx\right)^{\frac{1}{2}} \leqq A_{\sigma,\epsilon}(1 + | t |)^{(\sigma-\frac{1}{2})(1+\epsilon)}\left(\int_{-\infty}^{+\infty} | F(x) |^2 \, | x |^{2\sigma-1} \, dx\right)^{\frac{1}{2}}.$$

Our lemma is therefore proved.

Remark. We observe that the above proof yields the inequality

$$A_{\sigma,\epsilon} \leqq A_{\sigma_0}{}^{\theta}.$$

A simple argument then allows us to deduce the following fact: The constant A_σ which appears in (6.11) may be taken to be uniformly bounded in every closed subinterval of σ lying in $0 < \sigma < 1$.

This observation will be of use later.

CHAPTER II. UNIFORMLY BOUNDED REPRESENTATIONS.

7. Proofs of Theorem 5 and Theorem 6. Before presenting the details of the argument, we shall briefly discuss the main steps involved in the construction of the representations $g \to U^{\pm}(g, s)$.

Our representations are constructed on the space $\mathfrak{H} = \mathfrak{H}_{\frac{1}{2}}$ from representations $g \to V^{\pm}(g, s)$ on $\mathfrak{H}_\sigma$,[12] $s = \sigma + it$. The operators $U^{\pm}(g, s)$ and $V^{\pm}(g, s)$ are related by

[12] For the definition of the Hilbert space $\mathfrak{H}_\sigma$ see (6.1).

$$(7.1) \qquad U^{\pm}(g,s) = W(s,\tfrac{1}{2})V^{\pm}(g,s)W(\tfrac{1}{2},s),$$

where $W(s,\frac{1}{2})$ is the unitary transformation (6.2) of $\mathfrak{H}_\sigma$ onto $\mathfrak{H}_{\frac{1}{2}}$. The representations $g \to V^{\pm}(g,\frac{1}{2}+it)$ are obtained by simply transferring the representations $g \to v^{\pm}(g,\frac{1}{2}+it)$ of the continuous principal series from L_2 to $\mathfrak{H}_{\frac{1}{2}}$, by means of the Fourier transform. We also obtain the operators $V^{\pm}(g,s)$, $0 < Rs < \frac{1}{2}$, via the Fourier transform in a similar, but technically more involved, fashion from the representations $g \to v^{\pm}(g,s)$. To define the operators $V^{\pm}(g,s)$ for $\frac{1}{2} < R(s) < 1$ it is convenient to extend the notation $\sigma' = 1 - \sigma$ to complex s with $0 < Rs < 1$ by setting $s' = 1 - \bar{s}$; the transformation $s \to s'$ is then simply reflection about the line $\sigma = \frac{1}{2}$. Now the representation corresponding to an s with $\frac{1}{2} < Rs < 1$ is defined to be the contragredient of the representation corresponding to s'. Thus we put[18]

$$(7.2) \qquad V^{\pm}(g,s) = [V^{\pm}(g^{-1},s')]', \qquad \tfrac{1}{2} < Rs < 1.$$

It follows that

$$(7.3) \qquad [V^{\pm}(g^{-1},s)]' = V^{\pm}(g,s'), \qquad 0 < Rs < 1.$$

It will be shown in the course of the proof that the apparently arbitrary definition (7.2) is the natural one to make.

As a first step in the proof we shall establish the following lemma.

LEMMA 17. *The multipliers* $\phi^{\pm}$ *given by* (5.2), (5.3) *satisfy*

a) $\quad \phi^{\pm}(g,x,s) = \overline{\phi^{\pm}(g,x,\bar{s})},$

b) $\quad \phi^{\pm}(g_1g_2,x,s) = \phi^{\pm}(g_1,x,s)\phi^{\pm}(g_2,g_1x,s),$

c) $\quad \phi^{\pm}(g,g^{-1}x,s)\,dg^{-1}(x)/dx = \phi^{\pm}(g^{-1},x,1-s).$

The first relation, a) is immediate, b) is essentially a consequence of the chain rule for derivatives applied to (5.1), and c) follows by simple computations from b) upon setting $g_2 = g$ and $g_1 = g^{-1}$.

As the following lemma shows, it is natural to restrict s so that $0 \leqq Rx < 1$.

LEMMA 18. *Suppose* $s = \sigma + it$, *where* $0 \leqq \sigma \leqq 1$. *Then for each* $g \in G$, *the operators* $v^{\pm}(g,s)$ *are isometric on* L_p, *where* $p = (1-\sigma)^{-1}$.

Since the case $p = \infty$ is easily verified, we shall suppose $1 \leqq p < \infty$. Making the transformation $x \to g(x)$ we find that

---

[18] If A is an operator on $\mathfrak{H}_\sigma$, A' is the operator on $\mathfrak{H}_{\sigma'}$ given by (6.6) and (6.7).

$$\int_{-\infty}^{\infty} | f(x) |^p \, dx = \int_{-\infty}^{\infty} | bx + d |^{-2} | f(g(x)) |^p \, dx.$$

Now $| v^{\pm}(g,s) f(x) |^p = | bx + d |^{(2\sigma-2)p} | f(g(x)) |^p$, and since $(2\sigma - 2)p = -2$, it follows that $\| v^{\pm}(g,s) f \|_p = \| f \|_p$.

It is interesting to observe that when $p = (1-\sigma)^{-1}$ its conjugate index p' is given by $p' = (1-\sigma')^{-1}$. Thus the operators $v^{\pm}(g,s)$ and $v^{\pm}(g,s')$ give rise to a pair of isometric representations of G on L_p, $L_{p'}$ where $p = (1-\sigma)^{-1}$ and $s = \sigma + it$. Moveover, as the following lemma shows, these representations are contragredient.

LEMMA 19. *Let $s = \sigma + it$ and $p = (1-\sigma)^{-1}$, $0 < \sigma < 1$. Then for any $g \in G$, $f \in L_p$, and $h \in L_{p'}$,*

$$(v^{\pm}(g,s) f, h) = (f, v^{\pm}(g^{-1}, s') h). \tag{7.4}$$

To prove this we make the transformation $x \to g^{-1}(x)$ and find that

$$(v^{\pm}(g,s) f, h) = \int_{-\infty}^{\infty} \phi^{\pm}(g,x,s) f(g(x)) \overline{h(x)} \, dx$$

$$= \int_{-\infty}^{\infty} \phi^{\pm}(g, g^{-1}x, s) f(x) \overline{h(g^{-1}x)} \, (dg^{-1}(x)/dx) \, dx.$$

Thus, by c) of Lemma 17,

$$(v^{\pm}(g,s) f, h) = \int_{-\infty}^{\infty} f(x) \phi^{\pm}(g^{-1}, x, 1-s) \overline{h(g^{-1}x)} \, dx,$$

and now part a) of the same lemma shows that

$$(v^{\pm}(g,s) f, h) = (f, v^{\pm}(g^{-1}, s') h).$$

We now consider the representation spaces H_σ of the complementary series. These spaces are described in the following lemma.[14]

LEMMA 20. *Let $0 < \sigma < \frac{1}{2}$ and $p = (1-\sigma)^{-1}$. Then the inner product (5.5) is well defined for f in L_p, and the completion H_σ of L_p with respect to the norm $\| f \|_\sigma^2 = (f,f)_\sigma$ is unitarily equivalent to $\mathfrak{H}_\sigma$ via a mapping which coincides with the Fourier transform on L_p.*

To prove this, suppose first that $f \in L_1 \cap L_2$ and that F is its Fourier transform. By (6.14), which is valid for $0 < \epsilon < 1$, and the dominated convergence theorem we obtain

$$\int_{-\infty}^{\infty} | F(x) |^2 | x |^{2\sigma-1} \, dx = a_\sigma \int_{-\infty}^{\infty} f^* * f(x) | x |^{-2\sigma} \, dx. \tag{7.5}$$

[14] Lemmas 20, 21, and 22 are essentially restatements of known facts.

By Lemma 13 the left side of (7.5) is finite for $f \in L_p$, and by simple approximation arguments, it follows that the right side of (7.5) exists and equals the left side for all f in L_p. This shows that the formula (5.5) defines an inner product on L_p. Now observe that the Fourier transform of L_p, $1 < p \leqq 2$, includes the characteristics functions of finite intervals and their linear combinations. This observation together with (7.6) establishes the final statement of the lemma and concludes the proof.

As a consequence of Lemma 18 and Lemma 20, we obtain the fact that the representations $g \to v^{\pm}(g,s)$ are defined on a dense linear subset of H_σ, $0 < \sigma < \frac{1}{2}$. Moreover, as the following lemma shows, the operators $v^{\pm}(g,\sigma)$ extend uniquely to unitary operators on H_σ.

LEMMA 21. *Let $0 < \sigma < \frac{1}{2}$ and $p = (1-\sigma)^{-1}$. Then for f in L_p,*

$$\| v^{\pm}(g,\sigma) f \|_\sigma = \| f \|_\sigma. \tag{7.6}$$

In proving this, we use the fact that

$$g(x) - g(y) = (x-y)(bx+d)^{-1}(by+d)^{-1},$$

which follows by straightforward computation. Then making the transformations $x \to g(x)$ and $y \to g(y)$ we see that

$$\begin{aligned} \| f \|_\sigma^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\overline{f(y)} \, | x-y |^{-2\sigma} \, dx \, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(g(x))\overline{f(g(y))} \, | x-y |^{-2\sigma} \, | bx+d |^{2\sigma-2} \, | by+d |^{2\sigma-2} \, dx \, dy \\ &= \| v^{\pm}(g,\sigma) f \|_\sigma^2. \end{aligned}$$

Next we shall show that there exists a uniform bound independent of g for the operators $v^{\pm}(g,s)$ in H_σ; $s = \sigma + it$, $0 < \sigma < \frac{1}{2}$. In doing this, we consider the lower triangular subgroup of G consisting of elements g of the form

$$g = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}, \quad a \neq 0.$$

We make essential use of the fact that there are only two distinct double cosets of G modulo this subgroup. To be explicit, we introduce the group element

$$j = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{7.7}$$

and prove the following result.

LEMMA 22. *If $g \in G$ and is not lower triangular, there exist lower triangular group elements g_1 and g_2 such that*

$$g = g_1 j g_2. \tag{7.8}$$

We prove this by exhibiting such a decomposition. If $g \in G$ and is not lower triangular we may write

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad b \neq 0.$$

Then as is easily verified

$$g = \begin{bmatrix} 1 & 0 \\ db^{-1} & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} b^{-1} & 0 \\ a & b \end{bmatrix}.$$

In view of this result and the fact that $v^{\pm}(g_1 g_2, s) = v^{\pm}(g_1, s) v^{\pm}(g_2, s)$ for all g_1, g_2 in G, it is natural to consider the operators, $v^{\pm}(g, s)$, first for g in the lower triangular subgroup and then for $g = j$.

LEMMA 23. *Let $s = \sigma + it$, where $0 < \sigma < \frac{1}{2}$ and $-\infty < t < \infty$. Then*

1) *if g is lower triangular, $v^{\pm}(g, s)$ has a unique unitary extension to all of H_σ, and*[15]

2) $\| v^{\pm}(j, s) \|_\sigma \leqq A_\sigma (1 + |t|)^{\frac{1}{2}}$.

In proving 1), we suppose that $g = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}$, $a \neq 0$. Then by definition

$$v^{+}(g, s) : f(x) \to |a|^{2s-2} f(a^{-2}x + a^{-1}c), \tag{7.9}$$

and $v^{-}(g, s) = \operatorname{sgn}(a) v^{+}(g, s)$. Furthermore, $v^{+}(g, s) = |a|^{2it} v^{+}(g, \sigma)$; these relations together with Lemma 21 establish part 1). Turning now to the operators $v^{\pm}(j, s)$ we find that

$$v^{+}(j, s) : f(x) \to |x|^{2s-2} f(-1/x), \tag{7.10}$$

$$v^{-}(j, s) : f(x) \to \operatorname{sgn}(x) |x|^{2s-2} f(-1/x). \tag{7.11}$$

Now with the aid of the operators m_t^{+} and m_t^{-} given by (6.9) and (6.10) we can write $v^{+}(j, s) = m_t^{+} v^{+}(j, \sigma)$ and $v^{-}(j, s) = m_t^{-} v^{+}(j, \sigma)$. Since $v^{+}(j, \sigma)$ has a unitary extension, it follows that the bounds of the operators $v^{\pm}(j, s)$ are exactly the bounds of m_t^{+}, m_t^{-}, considered as acting in H_σ. Now using the fact that H_σ is unitarily equivalent to $\mathfrak{H}_\sigma$ and the definitions of M_t^{+}, M_t^{-} we obtain 2) as a consequence of Lemma 15.

---

[15] The symbol $\| \cdot \|_\sigma$ designates the bound of the operator on $\mathfrak{H}_\sigma$.

Finally, using Lemma 22 and Lemma 23, we find that

$$(7.12)\qquad \sup_g \| v^{\pm}(g,s) \|_\sigma \leqq A_\sigma(1+|t|)^{\frac{1}{2}}, \qquad 0<\sigma<\tfrac{1}{2}.$$

Because of (7.12) we may, and shall from now on, assume that the operators $v^{\pm}(g,s)$ are everywhere defined on H_σ.

Since H_σ and $\mathfrak{H}_\sigma$ are unitarily equivalent we may transfer the representations $g \to v^{\pm}(g,s)$ to $\mathfrak{H}_\sigma$ and obtain equivalent representations $g \to V^{\pm}(g,s)$. The operators $V^{\pm}(g,s)$ are obtained as follows: Let $\mathfrak{F}_\sigma$, $0<\sigma<\frac{1}{2}$, be the unitary transformation from H_σ to $\mathfrak{H}_\sigma$ that coincides with the Fourier transform (6.6) on L_p, $p=(1-\sigma)^{-1}$. In addition let $\mathfrak{F}_2$ be the Fourier transform restricted to L_2; we note that $\mathfrak{F}_2$ is unitary between L_2 and $\mathfrak{H}_{\frac{1}{2}}$. We now define $V^{\pm}(g,s)$ for $s=\sigma+it$ by

$$(7.13)\qquad V^{\pm}(g,s)=\mathfrak{F}_\sigma v^{\pm}(g,s)\mathfrak{F}_\sigma^{-1}, \qquad 0<\sigma\leqq\tfrac{1}{2}.$$

From (7.12) and the definitions (7.13), (7.2) we obtain the bounds

$$(7.14)\qquad \operatorname*{Sup}_g \| V^{\pm}(g,s) \|_\sigma \leqq A_\sigma(1+|t|)^{\frac{1}{2}}, \qquad 0<\sigma<1.$$

This result together with (7.1) implies

$$(7.15)\qquad \operatorname*{Sup}_g \| U^{\pm}(g,s) \|_\infty \leqq A_\sigma(1+|t|)^{\frac{1}{2}}, \qquad 0<\sigma<1.$$

Moreover, as the remark at the end of § 6 states, we may assume A_σ is bounded on any closed subinterval of $(0,1)$. Hence we have proved 5) of Theorem 5, and conclusions 2) and 3) follow from (7.13).

To show that (7.2) is a natural definition we consider once again the class of functions $\mathfrak{D}$ introduced in § 6. Recall that $F\in\mathfrak{D}$ if F is C^∞ and vanishes in a neighborhood of zero and outside a compact set.

LEMMA 24. *Suppose $F,H\in\mathfrak{D}$ and that f,h are their Fourier transforms. Then for all s in the strip $0<Rs<1$,*

$$(7.16)\qquad (v^{\pm}(g,s)f,h)=(V^{\pm}(g,s)F,H).$$

To prove this we suppose first of all that $0<Rs\leqq\frac{1}{2}$. Then $f, v^{\pm}(g,s)f\in L_p$, $p=(1-\sigma)^{-1}$, and $1<p\leqq 2$. Our result, (7.6), now follows from the definition of $V^{\pm}(g,s)F$ and the Parseval formula for $L_p, L_{p'}$, which is stated in Lemma 12. In case $\frac{1}{2}<Rs<1$, $V^{\pm}(g,s)=[V^{\pm}(g^{-1},s')]'$. Thus

$$(V^{\pm}(g,s)F,H)=(F,V^{\pm}(g^{-1},s')H)$$

By the result just established,

$$(F,V^{\pm}(g^{-1},s')H)=(f,v^{\pm}(g^{-1},s')h).$$

Now applying Lemma 19, we see that

$$(f, v^{\pm}(g^{-1}, s')h) = (v^{\pm}(g, s)f, h).$$

Thus (7.16) also holds for $\frac{1}{2} < Rs < 1$, and hence for all s in $0 < Rs < 1$.

To prove that the representations $g \to U^{\pm}(g, s)$, defined by (7.1), are continuous, it suffices to prove that the representations $g \to V^{\pm}(g, s)$ are; and for this, it is sufficient by (7.2) to consider the case $0 < Rs \leqq \frac{1}{2}$. Now if f is continuous and has compact support, it may be shown that for bounded functions h,

$$\int_{-\infty}^{\infty} \phi^{\pm}(g, x, s) f(g(x)) h(x)\, dx \to \int_{-\infty}^{\infty} f(x) h(x)\, dx$$

as $g \to e$, e being the identity in G. Because the representations $g \to v^{\pm}(g, s)$, $0 < Rs \leqq \frac{1}{2}$, are uniformly bounded on H_σ this is sufficient to insure their continuity. Hence the equivalent representations $g \to V^{\pm}(g, s)$ are also continuous.

It remains to prove conclusion 4) which refers to the analyticity of the operators $U^{\pm}(g, s)$. For this purpose we prove a result which has some interest in its own right.

LEMMA 25. *If g is a lower triangular matrix in G the operators $U^{\pm}(g, s)$ are independent of s, $0 < Rs < 1$.*

Let

$$g = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}, \quad a \neq 0,$$

and choose $F \in \mathfrak{H}_\sigma$. It then follows from (7.9) and well known properties of the Fourier transform that

$$V^{\pm}(g, s) : F(x) \to e^{ixac} \,|\, a \,|^{2s} F(a^2 x). \tag{7.17}$$

We also obtain the relation $V^{-}(g, s) = \operatorname{sgn}(a) V^{+}(g, s)$. Starting now with $F \in \mathfrak{H}$, we have by definition that

$$U^{\pm}(g, s) F = W(s, \tfrac{1}{2}) V^{\pm}(g, s) W(\tfrac{1}{2}, s) F.$$

Hence by (7.17),

$$V^{+}(g, s) W(\tfrac{1}{2}, s) : F(x) \to e^{ixac} \,|\, a \,|^{2s} \,|\, a^2 x \,|^{\frac{1}{2}-s} F(a^2 x)$$

and applying $W(s, \frac{1}{2})$ we get

$$U^{+}(g, s) : F(x) \to e^{ixac} \,|\, a \,|\, F(a^2 x). \tag{7.18}$$

Similarly, we obtain the relation $U^{-}(g,s)=\operatorname{sgn}(a)U^{+}(g,s)$. Thus we have proved the lemma.

This result shows that the inner products $(U^{\pm}(g,s)\xi,\eta)$ are constant as functions of s, and hence analytic, for any fixed lower triangular $g\in G$. Now if g is not lower triangular, it has a decomposition $g=g_1jg_2$ of the type (7.8). Since

$$U^{\pm}(g,s)=U^{\pm}(g_1,s)U^{\pm}(j,s)U^{\pm}(g_2,s),$$

where $U^{\pm}(g_i,s)$, $i=1,2$, are independent of s and have bounded inverses, it is sufficient to show that $(U^{\pm}(j,s)\xi,\eta)$ is analytic in s for each pair of vectors ξ, η in $\mathfrak{H}$. Recall the uniform bound, (7.15) for the representations $g\to U^{\pm}(g,s)$. Since the constant A_σ which appears is bounded as a function of σ over any closed subinterval of $(0,1)$, it is sufficient to prove that $(U^{\pm}(j,s)\xi,\eta)$ is analytic for a dense collection of vectors in $\mathfrak{H}$. Choose this collection to be the set $\mathfrak{D}$ of functions which are C^∞ and vanish in a neighborhood of zero, and outside a compact set. Pick $\xi=F$ and $\eta=H$ in $\mathfrak{D}$. Let $F_s(x)=|x|^{\frac{1}{2}-s}F(x)$ and put

$$H_s(x)=|x|^{\sigma-it-\frac{1}{2}}H(x).$$

It is then easily verified that

$$(U^{\pm}(j,s)\xi,\eta)=(V^{\pm}(j,s)F_s,H_s).$$

Denote the Fourier transforms of F_s, H_s by f_s, h_s. Then as F_s, H_s belong to $\mathfrak{D}$, Lemma 24 applies, and we see that

$$(U^{\pm}(j,s)\xi,\eta)=(v^{\pm}(j,s)f_s,h_s).$$

Now using (7.10), (7.11) we obtain

$$(7.19)\qquad (U^{+}(j,s)\xi,\eta)=\int_{-\infty}^{\infty}|x|^{2s-2}f_s(-1/x)\overline{h_s(x)}\,dx.$$

$$(7.20)\qquad (U^{-}(j,s)\xi,\eta)=\int_{-\infty}^{\infty}\operatorname{sgn}(x)\,|x|^{2s-2}f_s(-1/x)\overline{h_s(x)}\,dx.$$

Since

$$f_s(x)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}e^{ixy}\,|y|^{\frac{1}{2}-s}F(y)\,dy,$$

and in view of the various restriction on F, we may conclude that $f_s(x)$ has the following properties: it is jointly continuous as a function of x and s; it is analytic in s for each fixed x; and if s is restricted to any compact subset of the strip, $0<Rs<1$, $f_s(x)$ decreases as $|x|\to\infty$ as fast as any negative power of $|x|$. Since

$$\bar{h}_s(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ixy} \mid y \mid^{s-\frac{1}{2}} \overline{H(y)}\, dy$$

it has the same properties. It is now a very straightforward matter that (7.19), (7.20) can be obtained as uniform limits of functions analytic in s. Hence the inner products $(U^{\pm}(j,s)\xi,\eta)$ are analytic in s. This concludes the proof of Theorem 5.

COROLLARY. *Conclusion* 5) *of Theorem* 5 *may be strengthened as follows. Given any* $\epsilon > 0$, *then*

$$\sup_g \| U^{\pm}(g,s) \|_{\infty} \leqq A_{\sigma,\epsilon}(1 + | t |)^{|\sigma-\frac{1}{2}|(1+\epsilon)}$$

for $s = \sigma + it$, $0 < \sigma < 1$.

In proving the theorem we made use of the estimate given by (6.11). If, however, we had used the estimate given by Lemma 16, we would have obtained the above.

We shall now Prove 1) of Thetorem 6, which asserts that the representations $U^{\pm}(\cdot,s)$ and $U^{\pm}(\cdot,s')$ are contragredient.

In order to do this, we first combine (7.3) and (6.6) to obtain

$$V^{\pm}(g,s') = W_{\sigma}V^{\pm}(g^{-1},s)^{*}W_{\sigma}^{-1}. \tag{7.21}$$

It then follows by definition that

$$U^{\pm}(g,s') = W(s',\tfrac{1}{2})W_{\sigma}V^{\pm}(g^{-1},s)^{*}W_{\sigma}^{-1}W(\tfrac{1}{2},s').$$

Using the definitions of $W(s',\frac{1}{2}), W_{\sigma}$ together with the fact that $s'-\frac{1}{2}+\sigma-\sigma' = s-\frac{1}{2}$ we find that

$$W(s',\tfrac{1}{2})W_{\sigma} = W(s,\tfrac{1}{2}).$$

Substituting into the above we obtain

$$U^{\pm}(g,s') = W(s,\tfrac{1}{2})V^{\pm}(g^{-1},s)^{*}W(\tfrac{1}{2},s),$$

which implies,

$$U^{\pm}(g,s') = U^{\pm}(g^{-1},s)^{*}. \tag{7.22}$$

Hence we have proved part 1).

The second statement of Theorem 6 is easily seen to follow from the fact that the representations $g \to U^{+}(g,\sigma)$ are unitary for $0 < \sigma < 1$.

In fact, suppose that $g \to U(g,s)$ are any representations of G on $\mathfrak{H}$ such that

$$U(g,s') = U(g^{-1},s)^{*}$$

and for which the inner products $(U(g,s)\xi,\eta)$ are analytic in s. Then

$$U(\cdot,s)=U(\cdot,1-s)$$

if and only if the representations $U(\cdot,\sigma)$ are unitary for each σ, $0<\sigma<1$.

To prove this, we observe that the condition $U(\cdot,s)=U(\cdot,1-s)$ is, by analyticity, equivalent to the condition $U(\cdot,\sigma)=U(\cdot,1-\sigma)$ for $0<\sigma<1$. On the other hand, $1-\sigma=\sigma'$ so that $U(g^{-1},\sigma)^*=U(g,1-\sigma)$. Hence the above is equivalent to the condition $U(g,\sigma)=U(g^{-1},\sigma)^*$, $g\in G$.

It is interesting to note that the representations $U^-(\cdot,s)$ do not satisfy 2). This is a reflection of the known fact that they are not unitary when $\sigma\neq\frac{1}{2}$. In order to prove this, it is sufficient to show that the representations $v^-(\cdot,\sigma)$, $0<\sigma<\frac{1}{2}$, are not unitary. Without going into detail, we remark that this is a consequence of the relation

$$\begin{aligned}(7.23)\qquad &\| v^-(j,\sigma)f\|_\sigma{}^2-\|f\|_\sigma{}^2\\ &=a_\sigma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\operatorname{sgn}(x)sgn(y)-1)f(x)f(y)\,|\,x-y\,|^{-2\sigma}\,dxdy,\end{aligned}$$

which is valid for all f in L_p, $p=(1-\sigma)^{-1}$.

We suppose now that S is a bounded operator with a bounded inverse such that

$$SU^-(\cdot,s)S^{-1}=U^-(\cdot,1-s).$$

Replacing s by $1-s$ and making simple calculations we find that

$$S^2U^-(\cdot,s)=U^-(\cdot,s)S^2.$$

We shall assume the known fact that the unitary representations $U^-(\cdot,\frac{1}{2}+it)$, $t\neq 0$, are irreducible. It then follows that S^2 is a scalar multiple of the identity. For lower triangular group elements g of the form

$$g=\begin{bmatrix}a^{-1} & 0\\ c & a\end{bmatrix},\quad a\neq 0,$$

we know that

$$U^-(g,s):F(x)\to e^{ixac}aF(a^2x)$$

and

$$SU^-(g,s)=U^-(g,s)S.$$

Setting $a=1$ and then setting $c=0$ we find that S is the operation of multiplicity by a function, say K, with the property that $K(x)=K(a^2x)$. Since S^2 is a scalar multiple of the identity, we obtain the additional relation $(K(x))^2=\text{const.}$, which implies $K(x)=\text{const.}$ or $K(x)=(\text{const.})\operatorname{sgn}(x)$.

As the first alternative holds if and only if $U^-(\cdot, \sigma)$ is unitary for $0 < \sigma < 1$ we conclude that $K(x) = (\text{const.})\operatorname{sgn}(x)$.

We shall now define S by

$$S: F(x) \to \operatorname{sgn}(x) F(x), \tag{7.24}$$

and prove that $SU^-(g, s)S^{-1} = U^-(g, 1-s)$ for all g in G. This may be shown directly for all g; however, such a proof does not exhibit the crux of the matter, which, as it turns out, is the relation

$$SU^-(j, \sigma)S^{-1} = U^-(j, 1-\sigma).$$

We therefore proceed along different lines and first of all recall that the operators $U^-(g, s)$ are independent of s for lower triangular group elements g. For such g, the above relation becomes

$$SU^-(g, s) = U^-(g, s)S.$$

To verify this, suppose

$$g = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}, \quad a \neq 0.$$

Then

$$SU^-(g, s): F(x) \to \operatorname{sgn}(x) e^{ixac} aF(a^2x)$$

and

$$U^-(g, s)S: F(x) \to e^{ixac} a \operatorname{sgn}(a^2x) F(a^2x).$$

In view of the decomposition (7.8) it is therefore seen to be sufficient to prove the relation

$$SU^-(j, s) = U^-(j, 1-s)S;$$

Moreover, by analyticity, it is sufficient to prove this for $s = \sigma$, $0 < \sigma < \frac{1}{2}$.

Now

$$SU^-(j, \sigma)S^{-1} = U^-(j, \sigma')$$

if and only if

$$W(\tfrac{1}{2}, \sigma')SU^-(j, \sigma)S^{-1}W(\sigma', \tfrac{1}{2}) = W(\tfrac{1}{2}, \sigma')U^-(j, \sigma')W(\sigma', \tfrac{1}{2}).$$

Thus using the fact that S commutes with $W(\frac{1}{2}, \sigma')$ we see that it is sufficient to prove

$$V^-(j, \sigma') = SW(\sigma, \sigma')V^-(j, \sigma)W(\sigma', \sigma)S^{-1}, \quad 0 < \sigma < \tfrac{1}{2}. \tag{7.25}$$

In proving (7.25) we use the following considerations. The operation $SW(\sigma, \sigma')$ is multiplication by $\operatorname{sgn}(x)|x|^{2\sigma-1}$. Going over to the Fourier transform, this corresponds to convolution by $b_\sigma/(2\pi)^{\frac{1}{2}} \operatorname{sgn}(x)|x|^{-2\sigma}$, where

$$b_\sigma = i/\pi\Gamma(2\sigma)\sin\pi\sigma.$$

(This fact may be established in the same way as (6.13) was; for further discussion see the proof of Lemma 15, § 6.)

Recalling the definition of $V^-(j,\sigma)$ in terms of the Fourier transform, it then suffices to prove the following: the operation of convolution by $b_\sigma/(2\pi)^{\frac{1}{2}}\operatorname{sgn}(x)|x|^{-2\sigma}$, followed by the operation $f(x)\to\operatorname{sgn}(x)|x|^{-2\sigma}f(-1/x)$ is equal to the operation $f(x)\to\operatorname{sgn}(x)|x|^{2\sigma-2}f(-1/x)$ followed by convolution with $b_\sigma/(2\pi)^{\frac{1}{2}}\operatorname{sgn}(x)|x|^{-2\sigma}$. This leads to the verification

$$\int_{-\infty}^{\infty}\operatorname{sgn}(x)\operatorname{sgn}(-1/x-y)|x|^{-2\sigma}|-1/x-y|^{-2\sigma}f(y)\,dy$$
$$=\int_{-\infty}^{\infty}\operatorname{sgn}(y)\operatorname{sgn}(y-x)|y|^{2\sigma-2}|x-y|^{-2\sigma}f(-1/y)\,dy.$$

That this holds may be checked by the obvious change of variables.

The argument above needs to be made precise. We therefore argue as follows.

In proving (7.25) it clearly suffices to show that

$$(V^-(j,\sigma')SW(\sigma,\sigma')F,H)=(SW(\sigma,\sigma')V^-(j,\sigma)(F),H)$$

for $F,H\in\mathfrak{D}$. Let f be the Fourier transform of F. Put

$$f_1(x)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}e^{ixy}\operatorname{sgn}(y)|y|^{2\sigma-1}F(y)\,dy.$$

Then by what has been said before,

$$f_1(x)=b_\sigma/(2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty}\operatorname{sgn}(x-y)|x-y|^{-2\sigma}f(y)\,dy.$$

We define h, and h_1 similarly; thus it follows that

$$h_1(x)=b_\sigma/(2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty}\operatorname{sgn}(x-y)|x-y|^{-2\sigma}h(y)\,dy.$$

Since

$$(SW(\sigma,\sigma')V^-(j,\sigma)(F),H)=(V^-(j,\sigma)(F),SW(\sigma,\sigma')H),$$

it suffices in view of Lemma 24 to show that

$$(v^-(j,\sigma')f_1,h)=(v^-(j,\sigma)f,h_1). \tag{7.26}$$

Now,

$$(v^-(j,\sigma)f,h_1)=\int_{-\infty}^{\infty}\operatorname{sgn}(y)|y|^{2\sigma-2}f(-1/y)\overline{h_1(y)}\,dy$$

$$=-b_\sigma/(2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\overline{h(x)}f(-1/y)\operatorname{sgn}(y)\operatorname{sgn}(y-x)|y|^{2\sigma-2}|x-y|^{-2\sigma}\,dy\,dx.$$

On the other hand,

$$(v^-(j,\sigma')f_1,h)$$
$$=b_\sigma/(2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(x)f(y)\operatorname{sgn}(x)\operatorname{sgn}(-1/x-y)|\,x\,|^{-2\sigma}|\,x-y\,|^{-2\sigma}\,dydx.$$

If we make the change of variables $y\to -1/y$ in the first double-integral, then it is easily verified that this first double-integral equals the second double integral. This proves (7.26) and concludes the proof of Theorem 6.

CHAPTER III. The Fourier-Laplace Transform on the Group.

8. Hausdorff-Young theorem for the group and certain of its implications. Let $f\in L_1(G)$, and let us define the Fourier transform of f on G as follows:

$$\mathfrak{F}^{\pm}(s)=U^{\pm}(f,s)=\int_G U^{\pm}(g,s)f(g)\,dg, \tag{8.1}$$
$$0<R(s)<1,\quad f\in L_1(G).$$

$U^{\pm}(\cdot,s)$ is the analytic family of representations which act on $\mathfrak{H}$, and which were studied in §§ 5, 6 and 7. Because for each fixed s, $0<R(s)<1$, $U^{\pm}(\cdot,s)$ is a uniformly bounded representation, the integral appearing in (8.1) is well defined. Moreover, if $\xi,\eta\in\mathfrak{H}$, then

$$(\mathfrak{F}^{\pm}(s)\xi,\eta)=\int_G(U^{\pm}(g,s)\xi,\eta)f(g)\,dg.$$

An application of Fubini's theorem, and the analyticity of $U^{\pm}(\cdot,s)$ shows that

$$\int_C(F^{\pm}(s)\xi,\eta)\,ds=0,$$

for any closed curve C in $0<R(s)<1$.

Thus the Fourier transform $\mathfrak{F}^{\pm}(s)$ is not only well-defined when $f\in L_1(G)$, and $0<R(s)<1$, but is also an analytic operator-valued function of s in that strip.

The results of this section show, in a very precise way, that one may obtain similar results for the Fourier transform of functions in $L_p(G)$, $1\leqq p<2$. These facts are contained in the following theorem together with its corollaries.[16]

THEOREM 7. *Let $1<p<2$, and q be its conjugate index $1/p+1/q=1$.*

[16] The norms $\|\cdot\|_q$, $1\leqq q\leqq\infty$, are those introduced in § 2. We recall that $\|\cdot\|_2$ is the "Hilbert-Schmidt" norm while $\|\cdot\|_\infty$ denotes the operator bound.

There exists a measure $d\mu_{q,\sigma}(t)$ *so that*

$$(8.2) \qquad \left(\int_{-\infty}^{\infty} \| \mathfrak{F}^{\pm}(\sigma+it) \|_q^q \, d\mu_{q,\sigma}(t)\right)^{1/q} \leqq \| f \|_p,$$

f simple, $s = \sigma + it$, and $1/q < \sigma < 1/p$. For the measure $d\mu_{q,\sigma}(t)$ we have the following estimate: Given any $\delta > 0$, then:

$$d\mu_{q,\sigma}(t) \geqq A_{q,\sigma,\delta}(1 + |t|)^{1-q|\sigma-\frac{1}{2}|-\delta} \, dt.$$

COROLLARY 1. *For each fixed p, $1 < p < 2$, there exists a σ_0, $1/q < \sigma_0 < \frac{1}{2}$, so that*

$$\left(\int_{-\infty}^{+\infty} \| \mathfrak{F}^{\pm}(\sigma+it) \|_q^q \, dt\right)^{1/q} \leqq A_{q,\sigma} \| f \|_p$$

whenever $\sigma_0 < \sigma < 1 - \sigma_0$, and f is simple.

COROLLARY 2. *For each p, $1 \leqq p < 2$,*

$$\sup_{-\infty < t < \infty} \| \mathfrak{F}^{\pm}(\tfrac{1}{2} + it) \|_\infty \leqq A_p \| f \|_p, \ f \text{ simple}.$$

COROLLARY 3. *For each p, $1 \leqq p < 2$, $1/q < R(s) < 1/p$, $s = \sigma + it$,*

$$\| \mathfrak{F}^{\pm}(s) \|_\infty \leqq A_{p,\sigma,t} \| f \|_p, \text{ if simple}.$$

COROLLARY 4. *The Fourier transform, initially defined for $f \in L_1 \cap L_p$, has a unique bounded extension to all of $L_p(G)$, $1 \leqq p < 2$, with the following property: $U^{\pm}(f, \cdot)$ is for each $f \in L_p(G)$ analytic in s, for $1/q < R(s) < 1/p$. Moreover, the extension satisfies* (8.2) *as well as the conditions of Corollaries* 1 *through* 3.

Remarks. A strict analogue of the classical Hausdorff-Young theorem would have been a result like (8.2), but only for $\sigma = \frac{1}{2}$. The above results show, however, that the same conclusion holds for a proper strip which contains the line $\sigma = \frac{1}{2}$ in its interior. This, together with the analyticity of $\mathfrak{F}^{\pm}$, has far-reaching consequences. Once (8.2) has been proved, the results of Corollaries 2, 3, and 4 follow by rather standard "Phragmen-Lindelöf" type arguments.

It is possible to obtain somewhat stronger versions of Corollaries 2 and 3 by replacing the $\|\cdot\|_\infty$ operator norm by the norm $\|\cdot\|_q$. Since these latter results do not seem to have any immediate applications, we have not bothered to give their proofs.

A complete Fourier analysis of an arbitrary function (in the class $L_2(G)$) necessitates together with the continuous principal series also the discrete principal series. The discussion of the discrete principal series is much simpler, and is taken up in the next section.

Proof of Theorem 7. Let us consider the case $\mathfrak{F}^+$, that of $\mathfrak{F}^-$ being entirely similar. On account of the corollary to Theorem 5 (see § 7) we may write down the following inequality:

$$\sup_{-\infty<t<\infty}(1+|t|)^{-|\sigma-\frac{1}{2}|(1+\epsilon)}\,\|\mathfrak{F}^+(\sigma+it)\|_\infty \leqq A_{\sigma,\epsilon}\|f\|_1, \tag{8.3}$$
$$0<\sigma<1, \text{ and } \epsilon>0.$$

This inequality follows from the above quoted corollary and the observation that

$$\|\mathfrak{F}^+(\sigma+it)\|_\infty \leqq \sup_g \|U^+(g,s)\|_\infty \|f\|_1.$$

We know that $U^+(\cdot,\frac{1}{2}+it)$ is unitarily equivalent to the representation $v^+(\cdot,\frac{1}{2}+it)$ of the continuous principal series. This series, however, is contained in the Plancherel formula (see § 5). Hence we may write down the following inequality:

$$\left(\int_{-\infty}^{\infty}\|\mathfrak{F}^+(\tfrac{1}{2}+it)\|_2^2\, t^2(1+|t|)^{-1}\,dt\right)^{\frac{1}{2}} \leqq A\,\|f\|_2. \tag{8.4}$$

Here we have used the semi-trivial observation that,

$$t^2(1+|t|)^{-1}\leqq At\tanh\pi t, \qquad -\infty<t<\infty.$$

We shall apply Theorem 3 to inequalities (8.3) and (8.4) above. We argue as follows. Assume that σ is given and $1/q<\sigma<1/p$. We assume first that $\sigma<\frac{1}{2}$. Let α be a fixed real number with $0<\alpha<\sigma<\frac{1}{2}$, but otherwise arbitrary. Rewrite (8.3) with α instead of σ. It becomes

$$\sup_{-\infty<t<\infty}|1+|t|)^{c}\,\|\mathfrak{F}^+(\alpha+it)\|_\infty \leqq A_{\alpha,\epsilon}\|f\|_1 \tag{8.5}$$
$$\text{with } c=(\alpha-\tfrac{1}{2})(1+\epsilon).$$

Our given p, $1<p<2$, determines a parameter τ, $0<\tau<1$, with

$$1/p=(1-\tau)+\tau/2=1-\tau/2, \text{ and } 1/q=\tau/2.$$

Now if $1/q<\sigma<\frac{1}{2}$, there always exists an α, $0<\alpha<\sigma$, so that

$$\sigma=\alpha(1-\tau)+\beta\tau, \quad (\beta=\tfrac{1}{2}). \tag{8.6}$$

The above relation determines α uniquely, which α we now fix. In applying Theorem 3 to (8.4) and (8.5) we make the following further identifications:

$$\left\{\begin{array}{l} c=(\alpha-\frac{1}{2})(1+\epsilon)\\ a=1\\ b=\frac{1}{2}.\end{array}\right. \tag{8.7}$$

Now the result of Theorem 3 is

$$(8.8)\qquad \left(\int_{-\infty}^{\infty} \| \mathfrak{F}^{+}(\sigma+it)\|_{q}^{q}\,(1+|t|)^{dq}\,dt\right)^{1/q} \leqq A_{\epsilon,\tau}\,\|f\|_{p}$$

whenever f is simple.

A straightforward calculation leads to

$$(8.9)\qquad dq = 1 - |\sigma - \tfrac{1}{2}|\, q(1+\epsilon), \quad (\epsilon > 0).$$

Now given any $\delta > 0$, we can choose an $\epsilon > 0$, small enough so that

$$(8.10)\qquad dq = 1 - |\sigma - \tfrac{1}{2}|\, q - \delta.$$

Substituting this value of dq in (8.8) proves (8.2), whenever f is simple. The consideration of the case $\frac{1}{2} < \sigma < 1/p$ is carried out in the same manner once one defines

$$\mathfrak{F}^{+}(s) = \mathfrak{F}^{+}(1-s).$$

The consideration of $\mathfrak{F}^{-}(s)$ is analogous to $\mathfrak{F}^{+}(s)$. This concludes the proof of Theorem 5.

Proof of Corollary 1. Consider the quantity $1 - q\,|\sigma - \frac{1}{2}| - \delta$. This is the exponent that occurs in the measure $d\mu_{q,\sigma}(t)$. Recall that δ was arbitrary, except $\delta > 0$. Notice that if q is fixed we can make the quantity non-negative by choosing δ small enough and σ sufficiently close to $\frac{1}{2}$. However σ is also restricted by $1/q < \sigma < 1/p$. Thus it is clear that we can realize the conditions of the corollary if we take

$$\sigma_0 = \max(1/q, \tfrac{1}{2} - 1/q).$$

Hence for this choice of σ_0, the corollary is proved.

The proofs of the other corollaries necessitate the following lemma which is along very classical lines.

LEMMA 26. *Let $\Phi(s)$ be a (numerical-valued) function analytic in an open region which contains the strip*

$$\alpha \leqq R(s) \leqq \beta, \qquad \alpha < \beta.$$

Suppose that for some $c \geqq 0$,

$$\sup_{\alpha \leqq \sigma \leqq \beta} |\Phi(\sigma + it)| = O(|t|^{c}), \text{ as } |t| \to \infty,$$

and furthermore, for some q, $q \geqq 1$,

$$\int_{-\infty}^{\infty} |\Phi(\alpha + it)|^{q}\,dt \leqq 1, \qquad \int_{-\infty}^{\infty} |\Phi(\beta + it)|^{q}\,dt \leqq 1.$$

Let $\alpha < \gamma < \beta$.

Conclusion:

$$\sup_{-\infty<t<\infty} |\Phi(\gamma+it)| \leqq A.$$

A depends on α, β, γ, and q, but does not otherwise depend on c or Φ.

Proof. Let p be the index conjugate to q, $1/p+1/p=1$. Choose ϕ to be a continuous function on $(-\infty,\infty)$ which vanishes outside a finite interval, and satisfies

$$\int_{-\infty}^{\infty} |\phi(t)|^p \, dt \leqq 1, \tag{8.11}$$

but let ϕ be arbitrary otherwise.

Define $\Phi_1(\sigma+it)$ by

$$\Phi_1(\sigma+it) = \int_{-\infty}^{\infty} \Phi(\sigma+it+it_1)\phi(t_1)\,dt_1, \qquad \alpha \leqq \sigma \leqq \beta.$$

Then it is easy to verify that $\Phi_1(s)$ is analytic in an open region which contains $\alpha \leqq R(s) \leqq \beta$; that

$$\sup_{\alpha\leqq\sigma\leqq\beta} |\Phi_1(\sigma+it)| = O(|t|^c), \text{ as } |t|\to\infty;$$

and in view of the assumptions on Φ and (8.11) that

$$\sup_{-\infty<t<\infty} |\Phi_1(\alpha+it)| \leqq 1, \text{ and } \sup_{-\infty<t<\infty} |\Phi_1(\beta+it)| \leqq 1.$$

We are now in a position to apply the classical Phragmen-Lindelöf principle to Φ_1.[17] The conclusion is that $|\Phi_1|$ is bounded by 1 in the entire strip $\alpha \leqq R(s) \leqq \beta$. In particular,

$$|\Phi_1(\sigma)| \leqq 1, \qquad \alpha\leqq\sigma\leqq\beta.$$

Going back to the definition of Φ_1, we obtain

$$\left|\int_{-\infty}^{\infty} \Phi(\sigma+it_1)\phi(t_1)\,dt_1\right| \leqq 1, \qquad \alpha\leqq\sigma\leqq\beta.$$

Considering the arbitrariness of ϕ (except for condition (8.11)) the converse of Hölder's inequality shows:

$$\int_{-\infty}^{\infty} |\Phi(\sigma+it)|^q \, dt \leqq 1, \qquad \text{if } \alpha\leqq\sigma\leqq\beta. \tag{8.12}$$

For functions which are analytic in a strip and satisfy a uniform estimate like (8.12) there is a known variant of Cauchy's integral formula. It is

[17] See e.g. Titchmarsh [22; p. 181].

$$\Phi(\gamma+it)$$

$$(8.13)\qquad = \frac{1}{2\pi}\int_{-\infty}^{\infty}(\Phi(\alpha+it_1)/(\alpha+it_1-\gamma-it))\,dt_1$$
$$-\frac{1}{2\pi}\int_{-\infty}^{\infty}(\Phi(\beta+it_1)/(\beta+it_1-\gamma-it))\,dt_1,$$
$$\alpha<\gamma<\beta.$$

In Paley and Wiener ([18] pp. 3-5), (8.13) is demonstrated under the assumption corresponding to $q=2$ in (8.12). However, the proof in the general case, $q\geqq 1$, is no different.

If one applies Hölder's inequality to each of the integrals in (8.13) one obtains:

$$(8.14)\qquad \sup_{-\infty<t<\infty}|\Phi(\gamma+it)|\leqq A_{\alpha\beta\gamma q},$$

where

$$A_{\alpha\beta\gamma q}=\frac{1}{2\pi}[(\int_{-\infty}^{\infty}dt/((\gamma-\alpha)^2+t^2)^{p/2})^{1/p}$$
$$+(\int_{-\infty}^{\infty}dt/((\gamma-\beta)^2+t^2)^{p/2})^{1/p}].$$
$$(1/p+1/q=1).$$

A simple calculation shows,

$$(8.15)\qquad A_{\alpha,\beta,\gamma,q}\leqq c[(\gamma-\alpha)^{-1/q}+(\beta-\gamma)^{-1/q}],$$

and with c some absolute constant. This concludes the proof of the lemma.

Proof of Corollary 2. Let us assume for simplicity that $\|f\|_p=1$.

Considering the index σ_0 defined in Corollary 1, choose $\sigma_0<\sigma_1<\frac{1}{2}$, and keep σ_1 fixed throughout the rest of this argument. Now by the choice of σ_1 (and the normalization imposed on f) we have

$$(8.16)\qquad \begin{aligned}&\int_{-\infty}^{\infty}\|\mathfrak{F}^{\pm}(\sigma_1+it)\|_q^q\,dt\leqq A^q,\\ &\int_{-\infty}^{\infty}\|\mathfrak{F}^{\pm}(1-\sigma_1+it)\|_q^q\,dt\leqq A^q,\end{aligned}$$

with A independent of f, for some appropriate A.

Choose ξ and η to be two vectors in $\mathfrak{H}$, subject to the restriction $\|\xi\|\leqq 1$, $\|\eta\|\leqq 1$, but otherwise arbitrary. Now

$$|(\mathfrak{F}^{\pm}(s)\xi,\eta)|\leqq\|\mathfrak{F}^{\pm}(s)\|_\infty\leqq\|\mathfrak{F}^{\pm}(s)\|_q.$$

Hence if we let

$$\Phi(s)=1/A(\mathfrak{F}^{\pm}(s)\xi,\eta),$$

then

$$\int_{-\infty}^{\infty} |\Phi(\sigma_1+it)|^q\,dt \leqq 1, \qquad \int_{-\infty}^{\infty} |\Phi(1-\sigma_1+it)|^q\,dt \leqq 1. \tag{8.17}$$

However $\Phi(s)$ is clearly analytic in an open region containing $\sigma_1 \leqq R(s) \leqq 1-\sigma_1$—it is analytic in $0 < R(s) < 1$. Moreover, it satisfies the growth condition specified in Lemma 11, with $c = \frac{1}{2}$. We also notice $\sigma_1 < \frac{1}{2} < 1-\sigma_1$.

We then conclude:

$$\sup_{-\infty<t<\infty} |\Phi(\tfrac{1}{2}+it)| \leqq A_{\sigma_1,q}.$$

Going back to our definition this leads to

$$\sup_{-\infty<t<\infty} |(\mathcal{F}^{\pm}(\tfrac{1}{2}+it)\xi,\eta)| \leqq B_q.$$

Notice that $A_{\sigma_1,q}$ and hence B_q is independent of ξ,η. Taking the sup over all ξ,η, $\|\xi\| \leqq 1$, $\|\eta\| \leqq 1$ we obtain

$$\sup_{-\infty<t<\infty} \|\mathcal{F}^{\pm}(\tfrac{1}{2}+it)\| \leqq B_q.$$

If we now drop the normalization $\|f\|_p = 1$, we obtain the conclusion of Corollary 2. This concludes the proof.

Proof of Corollary 3. The proof is similar to that of Corollary 2 but is somewhat more complicated.

Let f be a simple function. We use inequality (8.2) which we have already proved for such f. We fix some $\delta > 0$, and assume momentarily that $\|f\|_p = 1$. Let us call

$$\lambda = 1 - |\sigma_1 - \tfrac{1}{2}|\,q - \delta,$$

where we choose $1/q < \sigma_1 < \frac{1}{2}$.

Then (8.2) becomes

$$\int_{-\infty}^{\infty} \|\mathcal{F}^{\pm}(\sigma_1+it)\|_q{}^q(1+|t|)^\lambda\,dt \leqq (A_{q,\sigma_1,\lambda})^q.$$

Choose $\xi,\eta \in \mathcal{H}$, with $\|\xi\| \leqq 1$, $\|\eta\| \leqq 1$, and let

$$\Psi(s) = (\mathcal{F}^{\pm}(s),\xi,\eta).$$

Then

$$\Psi(\sigma+it)| \leqq \|(|\mathcal{F}^{\pm}(\sigma+it)\|_\infty \leqq \|\mathcal{F}^{\pm}(\sigma+it)\|_q.$$

The above then becomes

$$\int_{-\infty}^{\infty} |\Psi(\sigma_1+it)|^q(1+|t|)^\lambda\,dt \leqq (A_{q,\sigma_1,\lambda})^q.$$

Since the formula (8.2) is symmetric in σ_1 and $1-\sigma_1$, one also obtains

$$\int_{-\infty}^{\infty} |\Psi(1-\sigma_1+it)|^q(1+|t|)^\lambda\, dt \leqq (A_{q,\sigma_1,\lambda})^q.$$

We let

$$\Phi(s) = c_1(2+s)^{\lambda/q}\Psi(s).$$

If we choose c_1 as appropriate constant (depending on q, σ_1, and λ) then the above inequalities become

$$\int_{-\infty}^{\infty} |\Phi(\sigma_1+it)|^q\, dt \leqq 1, \text{ and } \int_{-\infty}^{\infty} |\Phi(1-\sigma_1+it)|^q\, dt \leqq 1.$$

Moreover it is an easy matter to verify that $\Phi(s)$ satisfies the growth condition specified in Lemma 11. We may thus conclude, (see (8.15)),

$$|\Phi(\sigma+it)| \leqq c_2[|\sigma-\sigma_1|^{-1/q}] + [|1+\sigma-\sigma_1|^{-1/q}],$$
$$\sigma_1 < \sigma < 1-\sigma_1.$$

Going back to the definitions of Φ and Ψ the above becomes

$$|2+s|^{\lambda/q}|\mathfrak{F}^{\pm}(s)\xi,\eta)| \leqq c_3[|\sigma-\sigma_1|^{-1/q}+|1-\sigma_1-\sigma|^{-1/q}],$$
$$s=\sigma+it, \qquad \sigma_1<\sigma<1-\sigma_1.$$

Notice that the right-hand side is independent of ξ, and η. If we remember that ξ and η are arbitrary except $\|\xi\|\leqq 1$, $\|\eta\|\leqq 1$, and we take the sup of the left-hand side, dropping the restriction $\|f\|_p=1$, we then obtain

(8.18) $$|2+s|^{\lambda/q}\|\mathfrak{F}^{\pm}(s)\| \leqq c_3[|\sigma-\sigma_1|^{-1/q}+|1-\sigma_1-\sigma|^{-1/q}]\|f\|_p,$$

where $s=\sigma+it$, $\sigma_1<\sigma<1-\sigma_1$, $1/q<\sigma_1<\frac{1}{2}$,

$$c_3=c_3(q,\sigma_1,\lambda).$$

Notice that this formula actually holds for every s in the open strip $1/q<R(s)<1/p$. In fact, for such an s, we need only choose a σ_1 so that

$$\sigma_1<\sigma<1-\sigma_1, \text{ and } 1/q<\sigma_1<\tfrac{1}{2}.$$

If we now fix our λ and s, it is clear that (8.18) implies Corollary 3.

Proof of Corollary 4. It is clear from Corollary 3, that whenever $1/q<R(s)<1/p$, $\mathfrak{F}^{\pm}(s)$ has a unique bounded extension to all of $L_p(G)$.

Inequality (8.18) shows that the bounds are uniform whenever s is restricted to a compact subset of $1/q<R(s)<1/p$. But we know that $\mathfrak{F}^{\pm}$ is analytic in the strip $0<R(s)<1$, when $f\in L_1(G)\cap L_p(G)$.

Hence a simple limiting argument also shows that $\mathfrak{F}^{\pm}$ is analytic in $1/q < R(s) < 1/p$, for each fixed $f \in L_p(G)$.

Other limiting arguments (which we will not give) show that the extension $\mathfrak{F}^{\pm}$ to all of L_p also satisfies the inequalities (8.2) and those contained in Corollaries 1, 2, and 3.

This concludes our discussion of Corollary 4.

9. The discrete series. We now intend to investigate the form of the Hausdorff-Young theorem for our group, so far as it involves the discrete series.

As contrasted with the case of the continuous series considered above, we do not concern ourselves with an analytic structure in the discrete series. This lack is mitigated by the fact that in the Plancherel formula for the group, elements of the discrete series occur with weights bounded away from zero.

We begin by proving the following theorem.

THEOREM 8. *Let* $1 \leqq p \leqq 2$, and $1/q + 1/p = 1$. *Then*

$$(9.1) \qquad \Big(\sum_{k=0}^{\infty}(k+\tfrac{1}{2})\,\| D^{+}(f,k)\|_q^q + (k+1)\,\| D^{-}(f,k)\|_q^q\Big)^{1/q} \leqq \| f \|_p$$

whenever $f \in L_1(G) \cap L_p(G)$.

Proof. We consider the measure space M, defined as follows: The points of M are the pairs $(k, \pm)$, where k runs over the non-negative integers, and the second component is either $+$ or $-$ as indicated.

On M we define the measure dm as follows: The point $(k, +)$ is assigned the measure $k + \frac{1}{2}$; the point $(k, -)$ is assigned the measure $k + 1$.

We let $\mathfrak{H}$ denote a separable infinite-dimensional Hilbert space. In accordance with the discussion of § 2 we consider functions from M to bounded operators on $\mathfrak{H}$. In view of the discreteness of M, all such functions are automatically measurable.

We now define a mapping from simple functions on G to operator valued functions on M. The mapping, which we denote by T, is given by

$$T : f \rightarrow F = \{F(k, \pm)\},$$
$$F(k, \pm) = D^{\pm}(f, k),$$

and with

$$D^{\pm}(f, k) = \int_G D^{\pm}(g, k) f(g)\, dg.$$

As explicitly given, the representations $D^{\pm}(\cdot, k)$ act on different Hilbert

spaces. However, since all separable infinite dimensional Hilbert spaces are unitarily equivalent, we may assume that we deal with appropriate unitarily equivalent representations, all of which act on our given $\mathfrak{H}$.

Using the definitions of § 2, (9.1) becomes

$$\| T(f) \|_p \leqq \| f \|_q. \tag{9.2}$$

This is what we must prove

Observe that by definition,

$$\| T(f) \|_2^2 = \sum_{k=0}^{\infty} (k + \tfrac{1}{2}) \| D^+(f,k) \|_2^2 + (k+1) \| D^-(f,k) \|_2^2.$$

Hence, in view of the Plancherel formula for G, (see § 5), we have

$$\| T(f) \|_2 \leqq \| f \|_2. \tag{9.3}$$

Notice also that

$$\| T(f) \|_\infty = \sup_{k,\pm} \| D^\pm(f,k) \|_\infty$$

while

$$\| D^\pm(f,k) \|_\infty \leqq \| f \|_1,$$

since $D^\pm(\cdot, k)$ is unitary. We therefore have,

$$\| T(f) \|_\infty \leqq \| f \|_1. \tag{9.4}$$

We now use the general interpolation theorem of § 3. In the present case the operator T is independent of z, and so *a fortiori* satisfies the conditions of analyticity and admissible growth.

We let $(p_0, q_0) = (2,2)$, and $(p_1, q_1) = (1, \infty)$. Then it is easily verified that $1/p + 1/q = 1$, and that we may choose any p, $1 \leqq p \leqq 2$, by an appropriate choice of τ, $0 \leqq \tau \leqq 1$.

It is apparent that in the present case $A_0(y) = 1$ because of (9.3), and also $A_1(y) = 1$ because of (9.4).

The result of Theorem 3 is

$$\| T(f) \|_q \leqq A_\tau \| f \|_p.$$

Since $A_0(y) = A_1(y) = 1$, it follows that $\log A_\tau = 0$, and hence $A_\tau = 1$. Thus we have demonstrated (9.2), and therefore (9.1), whenever f is a simple function.

The extension of the inequality to all $L_1(G) \cap L_p(G)$ follows by standard limiting arguments. This concludes the proof of the theorem.

The following corollary is basic for our applications of the above theorem.

COROLLARY. *The mapping $f \to D^{\pm}(f,k)$ has a unique extension to all of $L_p(G)$, and this extension satisfies the following:*

$$\sup_{k,\pm} \| D^{\pm}(f,k) \|_{\infty} \leqq 2^{1-1/p} \| f \|_p \tag{9.5}$$

whenever $1 \leqq p \leqq 2$.

Proof. We consider first the case when $f \in L_1 \cap L_p$. Using (9.1) we obtain

$$(k+\tfrac{1}{2}) \| D^{+}(f,k) \|_q{}^q \leqq \sum_{k=0}^{\infty} (k+\tfrac{1}{2}) \| D^{+}(f,k) \|_q{}^q \leqq \| f \|_p{}^q.$$

Hence,

$$\| D^{+}(f,k) \|_q{}^q \leqq 1/(k+\tfrac{1}{2}) \| f \|_p \leqq 2 \| f \|_p{}^q.$$

A similar argument for $D^{-}(f,k)$ shows that

$$\sup_{k,\pm} \| D^{\pm}(f,k) \|_q \leqq 2^{1/q} \| f \|_p = 2^{1-1/p} \| f \|_p.$$

Since the operator norms used above are non-increasing with q, (see (2.2)), we conclude (9.5), whenever $f \in L_1 \cap L_p$.

In view of the inequality just proved it follows that the mapping $f \to D^{\pm}(f,k)$ has a unique extension to L_p which again satisfies (9.5).

CHAPTER IV. APPLICATIONS.

10. Boundedness of convolution operator. We are now in a position to obtain an important application of the analysis of the previous sections.

We shall find it convenient to adopt a slight change in our notation. In this section letters $x, y, z, \cdots$ will denote elements of the group G, and $f, g, h, \cdots$ functions on the group.

We recall the operation of convolution of two functions f and g, defined as follows

$$(f * g)(x) = \int_G f(xy^{-1}) g(y)\, dy,$$

$$dy = \text{Haar measure.}$$

Now if $f \in L_2$, and $g \in L_p$, $1 \leqq p \leqq 2$, then by Young's inequality (see [23]), $f * g$ is well defined and is in L_r, where $1/r = \frac{1}{2} + 1/p - 1$.

THEOREM 9. *Let $f \in L_2$, and $g \in L_p$, $1 \leqq p < 2$; if $h = f * g$, then $h \in L_2$, and*

$$\| h \|_2 \leqq A_p \| f \|_2 \| g \|_p, \tag{10.1}$$

where A_p does not depend on f or g. Hence the operation of convolution by a function $g \in L_p$, $1 \leqq p < 2$, is a bounded operator on L_2.

Remarks. Inequality (10.1) fails when $p = 2$. This is not surprising for many reasons; we indicate one such reason. Inequality (10.1) is essentially a statement of the fact that the Fourier transform of a function g in L_p, $1 \leqq p < 2$, is uniformly bounded. But a function in L_2 may be given by appropriately assigning its Fourier transform, and this may be done so that the Fourier transform is not uniformly bounded.

The statement which corresponds to (10.1) when G is, for example, a non-compact abelian group is false, as long as $p \neq 1$. This is so even in the case when G is the additive group of the real-line. We postpone further discussions of these matters to the next section.

Proof. It is sufficient to prove inequality (10.1) for a dense class of functions, and so we assume that f and g are in L_1 (in addition to the fact that f and g are respectively also in L_2 and L_p).

Notice that if $h = f * g$, and $x \to U_x$ is any (say unitary) representation, then

$$U_h = U_f \cdot U_g. \tag{10.2}$$

Here $U_f = \int_G f(x) U_x \, dx$, with similar definitions for U_g, and U_h.

Moreover by (2.13) and (10.2) we obtain

$$\| U_h \|_2 \leqq \| U_f \|_2 \, \| U_g \|_\infty. \tag{10.3}$$

We apply (10.3) successively to the cases when $U = U^{\pm}(\cdot, \tfrac{1}{2} + it)$, (the continuous principal series), and $U = D^{\pm}(\cdot, k)$, (the discrete series).

For the continuous principal series we apply Corollary 2 of Theorem 7, (with g in place of f); for the discrete series we similarly apply the corollary of Theorem 8. The result for the continuous series is

$$\| U^{\pm}(h, \tfrac{1}{2} + it) \|_2 \leqq A_p \, \| U^{\pm}(f, \tfrac{1}{2} + it) \|_2 \, \| g \|_p, \tag{10.4}$$

$1 \leqq p < 2$, with A_p independent of t.

The result for the discrete series is

$$\| D^{\pm}(h, k) \|_2 \leqq 2^{1-1/p} \, \| D^{\pm}(f, k) \|_2 \, \| g \|_p, \qquad 1 \leqq p \leqq 2. \tag{10.5}$$

Finally, we calculate $\| h \|_2$ and $\| f \|_2$ via the Plancherel formula, (see § 5).

It is to be noted that in computing the required norms, it makes no difference whether we use the representations $v^{\pm}(\cdot, \frac{1}{2}+it)$, or the unitarily equivalent representations $U^{\pm}(\cdot, \frac{1}{2}+it)$. Using (10.4) and (10.5) we then easily obtain

$$\| h \|_2 \leqq A_p \| f \|_2 \| g \|_p, \qquad 1 \leqq p < 2.$$

This proves (10.1), and hence the theorem.

From the above theorem, and with the use of various devices, it is possible to prove other inequalities like (10.1). All of these have in common the remarkable property that they hold for the group we are considering and also for compact groups, but fail in the simplest non-compact abelian instances. We shall limit ourselves to the proof of only one more such result.

COROLLARY. *Let $f \in L_2$, and $g \in L_2$. If $h = f * g$, then $h \in L_q$, $2 < q \leqq \infty$, and*

$$\| h \|_q \leqq A_q \| f \|_2 \| g \|_2, \tag{10.6}$$

where A_q does not depend on f or g.

Remark. By the results of the next section it will be seen that this corollary and the theorem from which it is derived are essentially equivalent results.

Proof. Let $k \in L_1 \cap L_p$, where $1/p + 1/q = 1$, but let k be arbitrary otherwise. Then,

$$\int_G h(x)k(x)\,dx = \int_G k(x) \int_G f(xy^{-1})g(y)\,dy\,dx$$
$$= \int_G g(y) \int_G k(x)f(xy^{-1})\,dx\,dy = \int_G g(y)l(y)\,dy,$$

where $l = \bar{f}^* * k$, with $f^*(x) = \bar{f}(x^{-1})$.

Hence,

$$\left| \int_G h(x)k(x)\,dx \right| = \left| \int_G g(y)l(y)\,dy \right| \leqq \| g \|_2 \| l \|_2.$$

However, by our theorem

$$\| l \|_2 \leqq A_p \| f \|_2 \| k \|_p,$$

since $1 \leqq p < 2$. Thus we have

$$\left| \int_G h(x)k(x)\,dx \right| \leqq A_p \| f \|_2 \| g \|_2 \| k \|_p.$$

Now take the sup of the left-hand side over k, such that $\| k \|_p = 1$. The result is

$$\| h \|_q \leqq A_p \| f \|_2 \| g \|_2,$$

and the corollary is proved.

11. Characterization of unitary representations of G. Let $g \to U_g$ be a unitary representation (not necessarily irreducible) on a Hilbert space $\mathfrak{H}$. We now introduce two notions which are basic for our characterization of the representations of G.

Definition. $\phi(g)$ is an *entry function,* if

$$\phi(g) = (U_g\xi, \eta), \qquad \xi, \eta \in \mathfrak{H}. \tag{11.1}$$

Definition. $g \to U_g$ is *extendable to* $L_p(G)$, if for some fixed p, $p \geqq 1$,

$$\| U_f \|_\infty \leqq A \| f \|_p, \text{ every } f \in L_1(G) \cap L_p(G), \tag{11.2}$$

with A independent of f.

It is interesting to note the following facts. Theorem 9, which dealt with the boundedness of the operation of convolution, can be restated by saying that the *regular* representation is extendable to $L_p(G)$ for every p, $1 \leqq p < 2$. We may further note that the corollary to Theorem 9 states that every entry function of the regular repersentation is in $L_q(G)$, for every $q > 2$.

As a preliminary matter, we obtain the following relation between the notions defined above.

LEMMA 27.[18] *The representation $g \to U_g$ is extendable to $L_p(G)$, if and only if for every pair $\xi, \eta \in \mathfrak{H}$, the entry function ϕ, defined in* (11.1), *lies in $L_q(G)$, where $1/p + 1/q = 1$.*

Assume first that U_g is extendable to L_p. Let $f \in L_1 \cap L_p$. Then

$$\int_G \phi(g) f(g)\, dg = \int (U_g\xi, \eta) f(g)\, dy$$
$$= \left(\int U_g f(g)\, dg\, \xi, \eta \right) = (U_f\xi, \eta).$$

Thus by (11.2)

$$\left| \int_G \phi(g) f(g)\, dg \right| \leqq |(U_f\xi, \eta)| \leqq A_p \| f \|_p \| \xi \| \| \eta \|.$$

[18] This lemma holds for any locally compact group.

We now limit ourselves to those f's for which $\| f \|_p = 1$, and we take the sup of the left-hand side. We then obtain $\| \phi \|_q \leqq A_p \| \xi \| \| \eta \|$, and thus $\phi \in L_q$. This proves the implication in one direction. To prove the converse we shall use the closed-graph theorem several times. We argue as follows.

For fixed η, consider the mapping

$$\xi \to (U_g \xi, \eta) = \phi(g)$$

as a mapping from $\mathfrak{H}$ to $L_q(G)$. By the assumptions of the lemma, it is clear that this mapping is everywhere defined on $\mathfrak{H}$; obviously it is linear. We next notice that it is closed. For supose that $\xi_n \to \xi$, and

$$\phi_n(g) = (U_g \xi_n, \eta) \to \phi_0(g) \text{ in } L_q \text{ norm.}$$

However, $\phi_n(g) \to \phi(g) = (U_g \xi, \eta)$, for every $g \in G$. Thus $\phi(g) = \phi_0(g)$ a. e., and $\phi_n(g) \to \phi(g)$ in L_q norm. This shows that the mapping is closed.

Hence,

$$\| \phi \|_q \leqq A_\eta \| \xi \|. \tag{11.3}$$

Similarly,

$$\| \phi \|_q \leqq B_\xi \| \eta \|. \tag{11.4}$$

Now let f be any function in $L_p(G)$. We propose to define U_f. We shall do this by defining $(U_f \xi, \eta)$, for every pair $\xi, \eta \in \mathfrak{H}$.

In fact set

$$(U_f \xi, \eta) = \int_G \phi(g) f(g) \, dg,$$

where $\phi(g) = (U_g \xi, \eta)$. Since $\phi \in L_q$, $f \in L_p$, and $1/p + 1/q = 1$, the integral is well-defined,, by Hölder's inequality. Hölder's inequality, (11.3), and (11.4) further show:

$$|(U_f \xi, \eta)| \leqq A_\eta \| \xi \| \| f \|_p, \tag{11.5}$$

and

$$|(U_f \xi, \eta)| \leqq B_\xi \| \eta \| \| f \|_p. \tag{11.6}$$

Now (11.6) shows that the vector $U_f \xi$ is well-defined for every $\xi \in \mathfrak{H}$. Moreover, (11.5) and a simple argument, prove that U_f is a closed operator. Hence, using the closed graph theorem, we obtain that U_f, for each $f \in L_p(G)$, is a bounded operator on $\mathfrak{H}$ (to itself).

Finally, consider the mapping

$$f \to U_f,$$

which is a mapping from $L_p(G)$ to $\mathfrak{B}$ = Banach space of bounded operators

on $\mathfrak{H}$ with usual norm. We have just seen that this mapping is everywhere defined. It is clear from the definition that this mapping is linear. We shall next see that it is closed. In fact, assume $f_n \to f$ in L_p norm, and that $U_{f_n} \to U_0$ in the operator norm. Then

$$(U_{f_n}\xi, \eta) \to (U_0\xi, \eta)$$

for every ξ and $\eta \in \mathfrak{H}$. By (11.5) it follows that

$$(U_{f_n}\xi, \eta) \to (U_f\xi, \eta).$$

Hence

$$(U_f\xi, \eta) = (U_0\xi, \eta).$$

Thus $U_f = U_0$. Therefore the mapping $f \to U_f$ is closed. A final application of the closed graph theorem gives

$$\| U_f \| \leqq A \| f \|_p.$$

This shows that $g \to U_g$ is extendable to L_p, and the lemema is completely proved.

We notice that the lemma proves that Theorem 9 and its corollary are equivalent propositions. It is to be observed that the identity representation (on the one-dimensional space) is *not* extendable to L_p if $p \neq 1$. Thus there are very simple representations which are not extendable to L_p if $p \neq 1$. We make one further remark before we proceed. Every entry function is automatically in $L_\infty(G)$. Hence a simple argument shows that if it is in $L_{q_0}(G)$, it is also in $L_q(G)$, where $q > q_0$. Therefore the lemma leads to the fact that if a representation is extendable to $L_{p_0}(G)$, and $p_0 > 1$, then it also is extendable to $L_p(G)$, for $1 \leqq p < p_0$.

We are now in a position to give our characterization of the irreducible unitary representations of the 2×2 real unimodular group G.

THEOREM 10. *Let $g \to U_g$ be an irreducible unitary representation of G. Assume U is not the identity representation. Then*

(a) *U is unitarily equivalent to an element of the discrete series if and only if U is extendable to $L_2(G)$.*

(b) *U is unitarily equivalent to an element of the continuous principal series if and only if U is extendable to every $L_p(G)$, $1 \leqq p < 2$, but is not extendable to $L_2(G)$.*

(c) *U is unitarily equivalent to the element of the complementary series corresponding to the parameter σ, $0 < \sigma < \frac{1}{2}$, if and only if U is*

extendable to $L_p(G)$ *for* $1 \leqq p < 1/(1-\sigma)$, *but is not extendable to* $L_{1/(1-\sigma)}(G)$.

COROLLARY. *Let* $g \to U_g$ *be an irreducible unitary representation different from the identity representation. Then* U *is unitarily equivalent to* (1) *an element of the discrete series,* (2) *an element of the continuous principal series, or,* (3) *the element of the complementary series corresponding to* σ, $0 < \sigma < \frac{1}{2}$, *if and only if respectively* (1′) *every entry function is in* $L_2(G)$, (2′) *every entry function is in* $L_q(G)$, $q > 2$, *but not every entry function is in* $L_2(G)$, *or,* (3′) *every entry function is in* $L_q(G)$, $q > 1/\sigma$, *but not every entry fuuction is in* $L_{1/\sigma}(G)$.

Before we pass to the proof of these facts we should like to clarify the difference of notation that we have adopted for the representations of G and that which is used in Bargmann's paper. The parameter σ, $0 < \sigma < \frac{1}{2}$, which we have used to identify the elements of the complementary series corresponds to Bargmann's parameter $\frac{1}{2} - \sigma$. There is also a difference in parametrization of the discrete series. We have called elements of the discrete series those which appear as discrete summands (with non-zero measure) in the Plancherel formula of the group. This exhausts Bargmann's discrete series, except for the representations which he labels $D^{+}{}_{1/2}$, and $D^{-}{}_{1/2}$. In our notation these two elements occur as follows. The representation $g \to U^{-}(g, \frac{1}{2})$ of the continuous principal series is not irreducible. It splits into the direct sum of $D^{+}{}_{1/2}$ and $D^{-}{}_{1/2}$. Thus in our notation we count $D^{+}{}_{1/2}$, and $D^{-}{}_{1/2}$ as elements of the continuous principal series. It is with these definitions in mind that the above theorem and corollary are stated.

Now to the proof. It is known that every irreducible unitary representation of the group is, except for the trivial representation, up to the unitary equivalence, either an element of the discrete series, the continuous principal series, or the complementary series.

By the corollary of Theorem 8, it follows that elements of the discrete series are extendable to $L_2(G)$. By Corollary 2 of Theorem 7 it follows that elements of the continuous principal series are extendable to $L_p(G)$, $1 \leqq p < 2$. If we consider the representation $g \to U^{-}(g, \frac{1}{2})$ we see that it is also extendable to L_p, for $1 \leqq p < 2$. However, this representation splits into two irreducible representations (which we have counted among the continuous principal series). A simple argument shows that each of these pieces is then also extendable to $L_p(G)$, $1 \leqq p < 2$.

Finally, Corollary 3 of Theorem 7 implies that the element corresponding to σ, $0 < \sigma < \frac{1}{2}$, is extendable to L_p, $1 \leqq p < 1/(1-\sigma)$. We must now show

that elements of the continuous principal series are not extendable to $L_2(G)$, and the element of the complementary series corresponding to σ is not extendable to $L_{1/(1-\sigma)}(G)$. We consider first the continuous principal series. By Lemma 27 it is sufficient to exhibit an entry function which is not in $L_2(G)$. We consider the parametrization of the group given by Bargmann with $0 < y < \infty$, $0 \leqq \mu \leqq 2\pi$, and $0 \leqq \nu < 2\pi$. In this case Haar measure becomes $(2\pi)^{-2}\, dy\, d\mu\, d\nu$, (see Bargmann (10.14)). We consider the "principal spherical function" corresponding to this representation. In Bargmann's notation this is $W_{00}(y)$, which has the asymptotic expansion, as $y \to \infty$,

$$W_{00}(y) \sim 2y^{-\frac{1}{2}} R(\beta_{00}(it, 0) y^{it}).$$

We also have

$$|\beta_{0,0}(it, 0)|^2 = (\coth \pi t)/4\pi t,$$

or $(\tanh \pi t)/4\pi t$, depending on whether we are dealing with $U^+(g, \frac{1}{2} + it)$ or $U^-(g, \frac{1}{2} + it)$, (see Bargmann (11.4), (11.7^a), and (11.7^b)). These asymptotic relations are valid except for $U^-(g, \frac{1}{2})$. Except for this case we can easily see that the element $W_{00}(y)$ is not in $L_2(0, \infty; dy)$, because of the factor $y^{-\frac{1}{2}}$. Thus the corresponding principal spherical functions are not in $L_2(G)$.

In considering the representation $U^-(g, \frac{1}{2})$ we recall that it splits into $D^+_{\frac{1}{2}}$, and $D^-_{\frac{1}{2}}$ (in Bargmann's notation). It is also demonstrated by Bargmann that the spherical functions corresponding to these representations are asymptotic to constant times $y^{-\frac{1}{2}}$. Thus these are also not in $L_2(G)$. The complementary series is dealt with similarly. Taking into account our difference in notation, we have $\beta_{00}(\frac{1}{2} - \sigma, 0) y^{-\sigma}$, as asymptotic expression (as $y \to \infty$) for the principal spherical function coresponding to σ, (see Bargmann (11.5)). Now clearly this function is not in $L_{1/\sigma}(0, \infty; dy)$. Hence the principal spherical function is not in $L_{1/\sigma}(G)$, and thus the representation is not extendable to $L_{1/(1-\sigma)}(G)$. If we recall Lemma 27, we see that Theorem 10 and its corollary are completely demonstrated.

We now pass to the consideration of not necessarily irreducible unitary representations. Let $g \to U_g$ be such a representation of our group on a separable Hilbert space $\mathfrak{H}$.

Using the von Neuman reduction theory [17], and following Segal [19], we may decompose the representation as follows.

The Hilbert space $\mathfrak{H}$ may be written as a direct integral $\int_{\oplus} \mathfrak{H}^\lambda\, d\sigma(\lambda)$ of Hilbert spaces $\mathfrak{H}^\lambda$. With respect to this decomposition, the representation $g \to U_g$ may be decomposed into $\{U^\lambda{}_g\}$, where $g \to U^\lambda{}_g$ is irreducible and unitary, for a. e. λ.

We do not wish to go into the background of these facts, or into the sense in which this reduction is unique. Aside from the simple manipulative facts which we shall use, we shall also use the following fact: Let A be an operator on $\mathfrak{H}$ which can be decomposed with respect to the above decomposition of $\mathfrak{H}$ into the direct integral of the $\mathfrak{H}^{\lambda}$'s. We write $A = \{A^{\lambda}\}$. Then $\| A \|_{\infty} = \operatorname{ess\,sup}_{\lambda} \| A^{\lambda} \|_{\infty}$.

Our theorem is the following: It may be viewed as an extension and clarification of Theorem 9 and its corollary.

THEOREM 11. *Let $g \to U_g$ be a unitary representation of G on $\mathfrak{H}$. Consider its reduction into a direct integral of irreducible unitary representations $g \to U^{\lambda}_g$. A necessary and sufficient condition that (except for a set of measure zero) every U^{λ}_g be unitarily equivalent to elements of the discrete or continuous principal series is that the representation $g \to U_g$ be extendable to $L_p(G)$ for every p, $1 \leqq p < 2$. Alternatively, the condition is equivalent with requiring that every entry function of the representation $g \to U_g$ be in every $L_q(G)$, $2 < q$.*

Proof. Assume first that, disregarding a set of measure zero, every U^{λ}_g is equivalent to either elements of the discrete series or of the continuous principal series. Let $f \in L_1(G) \cap L_p(G)$. Then

$$U_f = \{U^{\lambda}_f\}.$$

Now $\| U_f \|_{\infty} = \operatorname{ess\,sup}_{\lambda} \| U^{\lambda}_f \|_{\infty}$. Because of Corollary 2 of Theorem 7, and the corollary to Theorem 8, we obtain

$$\operatorname{ess\,sup}_{\lambda} \| U^{\lambda}_f \|_{\infty} \leqq A_p \| f \|_p, \qquad 1 \leqq p < 2;$$

we have disregarded the set of measure zero which does not correspond to either the discrete or continuous principal series. Hence,

$$\| U_f \| \leqq A_p \| f \|_p, \qquad 1 \leqq p < 2,$$

and $g \to U_g$ is extendable to every $L_p(G)$, $1 \leqq p < 2$. To prove the converse, we argue as follows. Let $\{f_n\}$ be a denumerable collection of functions on G which lie in every $L_p(G)$, and are dense in every $L_p(G)$, $1 \leqq p < \infty$. Now $U_{f_n} = U^{\lambda}_{f_n}$. Since $g \to U_g$ can be extended to $L_p(G)$, we have

$$\| U_{f_n} \|_{\infty} \leqq A_p \| f_n \|_p, \qquad 1 \leqq p < 2.$$

Thus,

$$\operatorname{ess\,sup}_{\lambda} \| U^{\lambda}_{f_n} \|_{\infty} \leqq A_p \| f_n \|_p, \qquad 1 \leqq p < 2.$$

Let E_n be the exceptional set of measure zero corresponding to the above ess sup.

Let $E = \bigcup E_n$; then E is still of measure zero. Now

$$\sup_{\lambda \notin E} \| U^\lambda{}_{f_n} \|_\infty \leqq A_p \| f_n \|_p, \qquad 1 \leqq p < 2.$$

Owing to the denseness of the collection $\{f_n\}$, we obtain

$$\| U^\lambda{}_f \|_\infty \leqq A_p \| f \|_p, \qquad 1 \leqq p < 2, \lambda \notin E, f \in L_1 \cap L_p.$$

Therefore $U^\lambda{}_g$ can be extended to L_p, $1 \leqq p < 2$, for every $\lambda \notin E$. By Theorem 10, and the fact that E is a set of measure zero we obtain that almost every $U^\lambda{}_g$ belongs to either the discrete or continouus principal series. Using Lemma 27 we obtain the alternate condition.

This concludes the proof of Theorem 11.

Let us now consider the additive group of the line. We shall show that the analogues of Theorem 9 fails for every $p \neq 1$, and that the analogue of the corollary of Theorem 9 fails if $q \neq \infty$.

In fact let $f(x) = (\log(2 + |x|))^{-1}$, $-\infty < x < \infty$. Then by the use of Theorem 124 of Titchmarsh [21] it may be shown f is the Fourier transform of a positive function which is in $L_1(-\infty, \infty)$. A simple application of the Plancherel theorem then shows that f is the convolution of two functions in $L_2(-\infty, \infty)$. However, clearly, $f \notin L_q(-\infty, \infty)$, if $q \neq \infty$. Thus the analogue of the corollary of Theorem 9 fails. Because of Lemma 27 applied to the regular representation on $L_2(-\infty, \infty)$, we see that the analogue of Theorem 9 fails if $p \neq 1$.

Let us consider the problem of whether Theorem 9 would hold for our group in the case $p = 2$. This, clearly, is equivalent to requiring that the regular representation of the group is extendable to $L_2(G)$. By an argument like that in the proof of Theorem 11, it would then follow that the regular representation can be written as a direct sum of representations equivalent to representations of the discrete series. This, of course, is not true.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] V. Bargmann, "Representations of the Lorentz group," *Annals of Mathematics*, vol. 48 (1947), pp. 568-640.

[2] S. Bochner, *Vorlesungen über Fouriersche Integral*, New York, 1948.

[3] A. P. Calderon and A. Zygmund, "On the theorem of Hausdorff-Young and its extension," *Contributions to Fourier Analysis*, Princeton, 1950.

[4] J. Dixmier, "Formes linéaires sur un anneau d'operateurs," *Bulletin de la Société Mathématique de France*, vol. 81 (1953), pp. 9-39.

[5] L. Ehrenpreis and F. I. Mautner, "Uniformly bounded representations of groups," *Proceedings of the National Academy of Sciences*, vol. 41 (1955), pp. 231-233.

[6] ———, "Some properties of the Fourier transform on semi-simple Lie groups, I," *Annals of Mathematics*, vol. 61 (1955), pp. 406-439.

[7] ———, "Some properties of the Fourier transform on semi-simple Lie groups, II," *Transactions of the American Mathematical Society*, vol. 84, No. 1 (1957), pp. 1-55.

[8] R. Godement, "Sur les relations d'orthogonalité de V. Bargmann I," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 225 (1947), pp. 521-523.

[9] ———, "Sur les relations d'orthogonalité de V. Bargmann II," *ibid.*, vol. 225 (1947), pp. 657-659.

[10] P. Halmos, *Measure Theory*, New York, 1950.

[11] G. H. Hardy and J. E. Littlewood, "Some more theorems concerning Fourier series and Fourier power series," *Duke Mathematical Journal*, vol. 2 (1936), pp. 355-381.

[12] ——— and G. Polya, *Inequalities*, Cambridge, 1934.

[13] Harish-Chandra, "Plancherel formula for the 2×2 real unimodular group," *Proceedings of the National Academy of Sciences*, vol. 38 (1952), pp. 337-342.

[14] ———, "Representations of semi-simple Lie groups VI," *American Journal of Mathematics*, vol. 78 (1956), pp. 564-628.

[15] E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Revised ed., American Mathematical Society Colloquium Publications, vol. XXXI, Providence, 1957.

[16] R. A. Kunze, "L_p Fourier transforms on locally compact unimodular groups," *Transactions of the American Mathematical Society*, vol. 89, No. 2 (1958), pp. 519-540.

[17] J. von-Neumann, "On rings of operators. Reduction theory," *Annals of Mathematics*, vol. 50 (1949, pp. 401-485.

[18] R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*, American Mathematical Society Colloquium Publications, vol. XIX, New York, 1934.

[19] I. E. Segal, *Decompositions of operator algebras. I*," Memoirs of the American Mathematical Society, No. 9, New York, 1951.

[20] E. M. Stein, "Interpolation of linear operators," *Transactions of the American Mathematical Society*, vol. 83 1956), pp. 482-492.

[21] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, Oxford, 1937.

[22] ———, *The theory of functions*, 2nd ed., Oxford, 1939.

[23] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Paris, 1938.

ON MAXIMAL FIRST ORDER PARTIAL DIFFERENTIAL OPERATORS.*

By H. O. Cordes.[1]

The modern spectral theory of selfadjoint differential operators, in all its various presentations, usually has the following starting point. One considers a linear formally selfadjoint differential operator L acting on either scalar or vector functions defined in some domain D. Then one defines a minimal and a maximal operator L_0 in $\mathfrak{D}(L_0)$ and L_1 in $\mathfrak{D}(L_1)$ to be the closure of the operator L in the domain of all sufficiently smooth functions vanishing in some neighborhood of the boundary and the (strict) adjoint of this operator respectively. Clearly L_0 is hermitian symmetric and is a restriction of L_1:

$$L_0 \subset L_1. \tag{1}$$

Both operators L_0 and L_1 are closed operators of a certain appropriate L_2-space into itself. Now, by von Neumann's Theory about extensions of closed hermitian symmetric operators, there exist selfadjoint extensions of L_0 (which also are restrictions of L_1) if any additional condition imposed on the operator L, as for instance reality or semiboundedness, guarantees the v. Neumann defect indices to be equal. Then, any such extension M being given, it is proved that the spectral representation

$$M = \int \lambda \, dE_\lambda \tag{2}$$

of M leads to an integral representation of arbitrary functions f in terms of the (regular and singular) eigenfunctions of M which are solutions of the equation

$$Lu = \lambda u. \tag{3}$$

This way, for instance, has been used by Stone in his book [18] for second order ordinary differential operators; then by Kodaira [11], Levinson [12], Levitan [13], Coddington [4] and many others for n-th order ordinary

* Received January 20, 1959.

[1] This paper has been prepared under the sponsorship of the Office of Naval Research, Contract No. Nonr 228(09).

operators; finally, for n-th order elliptic partial differential operators, by Garding [8], Vishik [19], [20], and Browder [1], and others. Now for ordinary differential operators a complete characterization of the possible operators M in $\mathfrak{D}(M)$ is easily done because then the defect indices of L_0 are finite and hence it easily can be shown that the domain $\mathfrak{D}(M)$ can be characterized by imposing boundary conditions on the functions of $\mathfrak{D}(L_1)$. These boundary conditions are conditions in the usual sense if the boundaries are regular. On the other hand, for a partial differential operator in a domain with regular boundary, the operator L_0 usually has infinite defect indices. This is why the characterization of the domain $\mathfrak{D}(M)$ by boundary conditions becomes a problem which in general has not been solved up to now.

However, if we understand the word "boundary condition" in some weakened manner, a solution can be achieved and this has ben done by several authors (Calkin [2], Vishik [19], [20], Hoermander [10], Phillips [14]). One then introduces a certain boundary space $\mathfrak{B}$ which is the quotient space

$$\mathfrak{G}(L_1)/\mathfrak{G}(L_0), \tag{4}$$

where $\mathfrak{G}(L_1)$ and $\mathfrak{G}(L_0)$ denote the graphs of the operators L_1 and L_0 respectively. It is easy to see that the residue class of some element $u \in \mathfrak{G}(L_1)$ with respect to the above quotient space depends only on the behavior of u at the boundary. Posing a generalized homogeneous boundary condition then simply means to impose on u that its residue class be in some given fixed subspace of $\mathfrak{B}$. One then can study the question whether or not a boundary condition is "well posed," i.e., leads to a selfadjoint (or more generally a maximal) operator M.

Nevertheless, it still remains the basic question to find out when such a generalized boundary condition can be posed in the usual way, i.e., imposing conditions on the function and its normal and tangential derivatives at the boundary only, and not in some neighborhood of the boundary. There are some more simple conditions, as for instance the Dirichlet condition, the von Neumann condition, and Hilbert's boundary condition of the third type, which have been investigated for certain types of differential operators. Also there are some nonlocal boundary conditions studied (Vishik [19], [20]), again for special types of differential operators.

The present paper deals with the discussion of a method which has many chances to work for very general classes of differential operators. We will discuss here maximality of formally selfadjoint first order partial differential operators acting on vector functions under boundary condition of a very large class which contains both local and nonlocal conditions. The conditions we

pose are boundary conditions in the strict sense mentioned above. It is interesting to see that ellipticity here does not play any further role.

In addition it seems to be quite obvious that similar arguments will work also in the case of higher order operators. The author is preparing another publication about the case of n-th order elliptic equations which will state quite analogous results.

The theory is based on investigating a relation between the inner product of the boundary space (which is essentially the inner product of the graph space) and another inner product which is defined for all "strips," i.e., all combinations of functions and sufficiently many normal derivatives at the boundary, under an ordinary L_2-norm with appropriate weight factor. The space completed under this norm will be called $\mathfrak{E}$. This relation will be established by a certain unbounded selfadjoint operator G in $\mathfrak{D}(G)$ of the boundary space $\mathfrak{B}$ into itself. The eigenvalue problem $G\,\phi = \lambda\phi$ is very closely related to some kind of eigenvalue problems with the parameter in the boundary condition studied first by Hilbert [9] for the operator

$$L = (\partial/\partial x)p(\partial/\partial x) + (\partial/\partial y)p(\partial/\partial y) \tag{5}$$

and generalized to the case of a second order elliptic operator on a Riemannian space by Sandgren [17].

After having this operator G established it is possible to study also boundary conditions which are imposed in the space $\mathfrak{E}$, i.e., which are ordinary boundary conditions.

The abstract classifications of the possible boundary conditions in the generalized sense for the case of the operators studied here has been investigated in great detail by R. S. Phillips [14]. Phillips studies not only selfadjoint opeartors but also a class of operators he calls maximal dissipative. A maximal dissipative extension M of L_0 simply is a maximal operator (in the v. Neumann sense) which essentially satisfies the condition

$$M + M^* \leqq 0. \tag{6}$$

This class of operators is important because a maximal dissiptative operator will be an infinitesimal generator of a semigroup of contractions. The selfadjoint case obviously arises if in (6) the equality sign holds. Then iM will be selfadjoint. Accordingly in this paper we shall look out for boundary conditions which will furnish maximal dissipative operators. Since the operator $-L$ instead of L also can be considered, it then also will be possible to find conditions which make M and $-M$ both maximal dissipative which amounts to iM being selfadjoint.

Finally it has to be mentioned that quite recently general locally dissipative conditions have been investigated by K. O. Friedrichs [7], P. D. Lax and R. S. Phillips [15]. This paper will essentially use the fact that a special operator M characterized by local boundary conditions is maximal dissipative. In other words we have to use the proofs mentioned above, which were achieved by use of mollifiers. The existence of the operator G mentioned above essentially follows from the well known result of K. O. Friedrichs [6] concerning identity of weak and strong solutions of first order systems and a certain continuation theorem proved in a concurrent paper of the author [5].

The main result is contained in Theorem 6.6. Section 2 contains the study of the operator G in $\mathfrak{D}(G)$ mentioned above. In sections 3 to 5 the two different boundary spaces are investigated and the tools for the proof of Theorem 6.6 are prepared.

The author wishes to express his apprecation for stimulation and for many valuable discussions about this subject which he had with Professor R. S. Phillips during the first period he worked on the paper. Especially he is indebted to him for the knowledge of an example which suggested the principal idea of this paper.

1. Auxiliary results. In this paragraph we will establish some known results which have to be used essentially in the following.

LEMMA 1.1. *Let*

$$A(s) = A(s_1, \cdots, s_p) = ((a_{ik}(s))) \tag{1.1}$$

be a symmetric $m \times m$*-matrix the coefficients* $a_{ik}(s)$ *of which are continuously differentiable with respect to* p *variables* $s = (s_1, \cdots, s_p)$ *in a certain domain* D_0. *Let* $P_0(s)$, $N_0(s)$, $Z_0(s)$ *be the orthogonal projections onto the subspaces of the* m*-component vector spaces which correspond to the eigenvalues* $\lambda_\nu(s) > 0$, $\lambda_\nu(s) < 0$, $\lambda_\nu(s) = 0$, *respectively.*

Then the coefficients of $P_0(s)$, $N_0(s)$ *and* $Z_0(s)$ *are bounded measurable functions and the coefficients of*

$$P(s) = A(s)P_0(s), \qquad N(s) = A(s)N_0(s) \tag{1.2}$$

are Lipschitz continuous in any compact subregion of D_0. *If*

$$u(s) = (u_1(s), u_2(s), \cdots, u_m(s)) \tag{1.3}$$

is any bounded measurable m*-component vector function defined in* D_0 *for which*

$$v(s) = A(s)u(s) \tag{1.4}$$

is Lipschitz continuous in any compact subregion of D_0, then the same holds for

$$u_+(s) = P_0(s)u(s), \qquad u_-(s) = N_0(s)u(s), \tag{1.5}$$

i.e., $u_+(s)$, $u_-(s)$ are bounded measurable in D_0 and $v_\pm(s) = A(s)u_\pm(s)$ is Lipschitz continuous in any compact subregion of D_0.

Proof. It suffices to prove every statement for $P_0(s)$, $P(s)$ and $v_+(s)$ only. Let D_1 be any compact subregion of D_0 and let σ be a positive number such that $|A(s)u| \leqq (\sigma-1)|u|$ holds for every m-component vector u and for every $s \in D_1$. Here $|u|^2 = \sum_{i=1}^{m} |u_i|^2$ with u_i being the components of u. Accordingly we denote $\sum_{i=1}^{m} \bar{u}_i v_i$ by $\bar{u}v$. If $R_z(s) = (A(s)-z)^{-1}$ denotes the resolvent of $A(s)$ and if for $0 \leqq \epsilon < \sigma$ C_ϵ denotes the closed path in the complex z-plane which is the boundary of the region $\epsilon \leqq |z| \leqq \sigma$, $\operatorname{Re} z \geqq 0$, then it is well known that

$$P_0(s) = \frac{1}{2\pi i} \lim_{\epsilon\to 0,\, \epsilon>0} \int_{\mathfrak{C}_\epsilon} R_z(s)\, dz. \tag{1.6}$$

The integral has to be taken as a Cauchy mean value in case $\mathfrak{C}_\epsilon$ crosses any eigenvalue of $A(s)$. On the other hand

$$\begin{aligned} &R_{z_1}(s^1) - R_{z_2}(s^2) \\ &\quad = R_{z_2}(s^2)(A(s^2) - z_2)R_{z_1}(s^1) - R_{z_2}(s^2)(A(s^1) - z_1)R_{z_1}(s^1) \\ &\quad = R_{z_2}(s^2)\{A(s^2) - A(s^1) + (z_1 - z_2)\}R_{z_1}(s^1). \end{aligned}$$

This shows that for every z_0 which is not an eigenvalue of $A(s^0)$ the coefficients of $R_z(s)$ are continuous and even continuously differentiable at (s^0, z_0). Since $A(s)$ is symmetric, all eigenvalues are real and obviously the path $\mathfrak{C}_\epsilon$ does not contain any eigenvalue except perhaps at $z = \epsilon$. Especially we get

$$(\partial/\partial s_i)R_z(s) = -R_z(s)((\partial/\partial s_i)A(s))R_z(s). \tag{1.7}$$

This shows that $P_0(s)$ is generated by a twice iterated limit process from a set of matrix functions $R_z(s)$ having continuously differentiable coefficients. Consequently $P_0(s)$ is measurable and, of course, bounded. Further it is well known that

$$P(s) = \frac{1}{2\pi i} \int_{\mathfrak{C}_0} R_z(s)A(s)\, dz \tag{1.8}$$

and

$$v_+(s) = \frac{1}{2\pi i} \int_{\mathfrak{C}_0} R_z(s)A(s)u(s)\, dz = \frac{1}{2\pi i} \int_{\mathfrak{C}_0} R_z(s)v(s)\, dz. \tag{1.9}$$

It suffices to prove the Lipschitz continuity of $v_+(s)$, since for $u(s)=u_0$ $=$ const. the Lipschitz continuity of $P(s)$ also follows.

For $s^1, s^2 \in D_1$ and $z \neq \lambda_\nu(s^1)$, $z \neq \lambda_\nu(s^2)$, $\nu = 1, \cdots, m$, we get

$$\begin{aligned} &R_z(s^1)v(s^1) - R_z(s^2)v(s^2) \\ (1.10) \qquad &= \{R_z(s^1) - R_z(s^2)\}v(s^1) + R_z(s^2)(v(s^1) - v(s^2)) \\ &= R_z(s^2)(A(s^2) - A(s^1))R_z(s^1)u(s^1) + R_z(s^2)(v(s^1) - v(s^2)). \end{aligned}$$

Hence

$$\begin{aligned} &| v_+(s^1) - v_+(s^2) | \\ (1.11) \qquad &\leqq \tfrac{1}{2\pi} \left| \int_{\mathfrak{C}_0} R_z(s^2)(A(s^2) - A(s^1))R_z(s^1)A(s^1)u(s^1)dz \right| \\ &\quad + \tfrac{1}{2\pi} \left| \int_{\mathfrak{C}_0} R_z(s^2)(v(s^1) - v(s^2))dz \right|. \end{aligned}$$

For the second term we immediately get

$$(1.12) \qquad \tfrac{1}{2\pi} \left| \int_{\mathfrak{C}_0} R_z(s^2)(v(s^1) - v(s^2)) \right| dz \leqq | v(s^1) - v(s^2) | \leqq c \, | s^1 - s^2 |,$$

where

$$(1.13) \qquad c = \sup_{s^1 \neq s^2;\, s^1, s^2 \in D_1} \{ | v(s^1) - v(s^2) | \, | s^1 - s^2 |^{-1} \}.$$

Here we used that $\frac{1}{2\pi} \int_{\mathfrak{C}_0} R_z(s^2)\, dz$ is a contraction matrix, as easily can be proved.

In order to estimate the first term let ϕ^ν, λ_ν, $\nu = 1, \cdots, m$, and ψ^ν, μ_ν, $\nu = 1, \cdots, m$, be the eigenvectors and eigenvalues of $A(s^1)$ and $A(s^2)$ respectively and denote

$$(1.14) \qquad \overline{\psi^\kappa}(A(s^2) - A(s^1))\phi^\nu = q_{\kappa\nu}.$$

Clearly $| q_{\nu\kappa} | \leqq c \, | s^1 - s^2 |$, $s^1, s^2 \in D_1$, where c is independent of s^1 and s^2. Obviously then we get

$$(1.15) \qquad \begin{aligned} R_z(s^1) A(s^1) u &= \sum_{\lambda_\nu \neq 0} \lambda_\nu (\lambda_\nu - z)^{-1} (\overline{\phi^\nu} u) \phi^\nu \\ R_z(s^2) u &= \sum_{\kappa=1}^{m} (\mu_\kappa - z)^{-1} (\overline{\psi^\kappa} u) \psi^\kappa \end{aligned}$$

and thus

$$\begin{aligned} &\tfrac{1}{2\pi} \left| \int_{\mathfrak{C}_0} R_z(s^2)(A(s^2) - A(s^1)) R_z(s^1) A(s^1) u(s^1)\, dz \right| \\ (1.16) \qquad &= \tfrac{1}{2\pi} \left| \sum_{\lambda_\nu \neq 0} \sum_{\kappa=1}^{m} q_{\kappa\nu} (\overline{\phi^\nu} u) \psi^\kappa \int_{\mathfrak{C}_0} \lambda_\nu [(\lambda_\nu - z)(\mu_\kappa - z)]^{-1} dz \right| \\ &\leqq \tfrac{1}{2\pi} \sum_{\lambda_\nu \neq 0} \sum_{\kappa=1}^{m} | q_{\kappa\nu} | \, | \overline{\phi^\nu} u | \left| \int_{\mathfrak{C}_0} \lambda_\nu [(\lambda_\nu - z)(\mu_\kappa - z)]^{-1} dz \right|. \end{aligned}$$

If $\lambda_\nu = \mu_\kappa \neq 0$ then the corresponding integral in the above right hand side vanishes. But for $\lambda_\nu \neq \mu_\kappa$,

$$\lambda_\nu[(\lambda_\nu - z)(\mu_\kappa - z)]^{-1} = \lambda_\nu(\mu_\kappa - \lambda_\nu)^{-1}[(\lambda_\nu - z)^{-1} - (\mu_\kappa - z)^{-1}]. \tag{1.17}$$

Consequently

$$\int_{\mathfrak{C}_0} \lambda_\nu[(\lambda_\nu - z)(\mu_\kappa - z)]^{-1}\, dz = 0 \tag{1.18}$$

if λ_ν and μ_κ are both positive or both negative. Further we get

$$\left| \int_{\mathfrak{C}_0} \lambda_\nu[(\lambda_\nu - z)(\mu_\kappa - z)]^{-1}\, dz \right| = 2\pi \left| \lambda_\nu(\mu_\kappa - \lambda_\nu)^{-1} \right| \tag{1.19}$$

if λ_ν and μ_κ have opposite sign and

$$\left| \int_{\mathfrak{C}_0} \lambda_\nu[(\lambda_\nu - z)(\mu_\kappa - z)]^{-1}\, dz \right| = \pi \left| \lambda_\nu(\mu_\kappa - \lambda_\nu)^{-1} \right| \tag{1.20}$$

if $\mu_\kappa = 0$. But in both cases we also get $|\lambda_\nu(\mu_\kappa - \lambda_\nu)^{-1}| \leqq 1$. Consequently

$$\begin{aligned} &\frac{1}{2\pi} \left| \int_{\mathfrak{C}_0} R_z(s^2)(A(s^2) - A(s^1))A(s^1)u(s^1)\, dz \right| \\ &\qquad \leqq \sum_{\lambda_\nu \neq 0} \sum_{\kappa=1}^{m} |q_{\kappa\nu}|\, |\overline{\phi^\nu u}| \leqq c\, |s^1 - s^2|. \end{aligned} \tag{1.21}$$

This proves the lemma.

2. A special eigenvalue problem. Let the operator L_1 in $\mathfrak{D}(L_1)$ be defined by

$$L_1 u = \sum_{i=1}^{n} a_i(x)\partial u/\partial x_i + b(x)u(x) \tag{2.1}$$

for $u \in \mathfrak{D}(L_1)$. The matrix functions $a_i(x)$, $b(x)$ are assumed to be $m \times m$-matrices defined and continuous in a domain D of $(x_1, \cdots, x_n)$-space and on its boundary Γ. The matrices $a_i(x)$ are assumed to be hermitian symmetric and with uniformly Hölder continuous first derivatives on $D + \Gamma$. Further the domain D is assumed to be bounded with its boundary Γ consisting of a finite number of simple nonintersecting hypersurfaces which are all twice Hölder continuously differentiable. We assume further the operator iL_1 to be formally selfadjoint, i. e.,

$$b(x) + b^*(x) = \sum_{i=1}^{n} (\partial/\partial x_i)(a_i(x)), \qquad x \in D + \Gamma, \tag{2.2}$$

where $b^*(x)$ denotes the adjoint of the matrix $b(x)$ (i. e., the transposed and complex conjugate matrix).

Let the domain $\mathfrak{D}(L_1)$ be the space of all complex valued m-componen vector functions defined in $D+\Gamma$ satisfying the following conditions:

a) u, $\partial u/\partial x_i$, $i=1,\cdots,n$, are continuous in D.

b) $u(x)$ is uniformly bounded in D.

c) $\lim_{\epsilon\to 0} u(x-\epsilon\nu)=u(x)$ holds for every $x\in\Gamma$, except possibly a $(n-1)$-dimensional null set.

d) $v(x)=A(x)u(x)$ is continuous on $D+\Gamma$ and Lipschitz continuou on Γ.

e) $\int_D |L_1u|^2\,dx<\infty$.

Let the $m\times m$-matrix $A(x)$ be continuously differentiable and let its firs derivatives satisfy a Hölder condition along the boundary Γ. Further let

$$A(x)=\sum_{i=1}^{n} a_i(x)\nu_i(x) \text{ on } \Gamma, \tag{2.3}$$

where $\nu=(\nu_1(x),\cdots,\nu_n(x))$ denotes the exterior normal on Γ.

Let $B(x)$ be the positive square root of $(A(x))^2$:

$$B(x)\geqq 0,\qquad (B(x))^2=(A(x))^2. \tag{2.4}$$

For the space $\mathfrak{D}(L_1)$ we introduce the following two bilinear forms:

$$[u,v]=\int_\Gamma \overline{u(x)}B(x)v(x)\,d\sigma, \tag{2.5}$$

$$(u,v)=\int_D (\overline{L_1u}L_1v+\bar{u}v)\,dx. \tag{2.6}$$

Here $d\sigma$ denotes the area element on Γ, and by $\bar{z}w$ we mean the local inne product of the two complex valued m-component vectors z and w:

$$\bar{z}w=\sum_{i=1}^{m}\bar{z}_iw_i,\qquad |z|^2=\sum_{i=1}^{m}|z_i|^2. \tag{2.7}$$

Clearly both forms are positive:

$$[u,u]\geqq 0,\qquad (u,u)\geqq 0,\qquad u\in\mathfrak{D}(L_1). \tag{2.8}$$

In addition the form (u,u) is positive definite: $(u,u)>0$, $u\neq 0$, $u\in\mathfrak{D}(L_1)$

By adding ideal elements we complete the domain $\mathfrak{D}(L_1)$ with respec to the positive definite form (u,u) to a Hilbert space which we call $\mathfrak{H}^0$. Sinc the metric

$$\|u\|=\{(u,u)\}^{\frac{1}{2}} \tag{2.9}$$

is stronger than that induced by the ordinary inner product in L_2

$$\langle u, v\rangle = \int_D \bar{u}v\,dx, \tag{2.10}$$

the space $\mathfrak{H}^0$ can be identified with a certain subspace of the space $\mathfrak{H}^0$ of all (classes of equivalent) measurable m-component vector functions square integrable over D:

$$\langle\langle u\rangle\rangle^2 = \int_D |u|^2\,dx < \infty. \tag{2.11}$$

So the ideal elements can be represented by square integrable functions too.

We still consider $\mathfrak{D}(L_1)$ as a subspace of $\mathfrak{H}$, since by definition it is the domain of the operator L_1, which is an operator of $\mathfrak{H}$ into itself. In order to discriminate between $\mathfrak{D}(L_1)$ considered as a subspace of $\mathfrak{H}^0$ and of $\mathfrak{H}$ we denote the set of all elements of $\mathfrak{D}(L_1)$ considered as a subspace of $\mathfrak{H}^0$ by $\mathfrak{H}^0(L_1)$. This additional notation will prove to be necessary in the later considerations.

Our first aim is to prove the following

THEOREM 2.1. *There is a selfadjoint operator G^0 defined in a dense subspace $\mathfrak{D}(G^0)$ of $\mathfrak{H}^0$ such that*

$$\mathfrak{D}(G^0) \supset \mathfrak{H}^0(L_1) \text{ and } [u, v] = (u, G^0 v), \qquad u, v \in \mathfrak{H}^0(L_1). \tag{2.12}$$

Proof. First note that by (2.2) and Green's formula the following relation holds:

$$\int_D (\bar{u}L_1 v + \overline{L_1 u}v)\,dx = \int_\Gamma \bar{u}Av\,d\sigma;\ u, v \in \mathfrak{D}(L_1). \tag{2.13}$$

We denote the bilinear form introduced in (2.13) by $Q(u, v)$. By Schwartz' inequality

$$|Q(u, v)| \leqq \|u\|\,\|v\|, \qquad u, v \in \mathfrak{D}(L_1). \tag{2.14}$$

Consequently there exists a uniquely determined bounded linear selfadjoint operator Q^0 defined in the whole space $\mathfrak{H}^0$ such that

$$Q(u, v) = (Q^0 u, v), \qquad u, v \in \mathfrak{H}^0(L_1). \tag{2.15}$$

Now for any $x \in \Gamma$ let

$$P(x) = \tfrac{1}{2}(B(x) + A(x)), \qquad N(x) = \tfrac{1}{2}(B(x) - A(x)), \tag{2.16}$$

with the matrices $A(x)$ and $B(x)$ defined in (2.3) and (2.4). Denote the spaces of all $v \in \mathfrak{H}^0(L_1)$ satisfying

$$P(x)v(x) = 0, \text{ or } N(x)v(x) = 0, \qquad x \in \Gamma, \tag{2.17}$$

by $\mathfrak{H}^0{}_-'$ and $\mathfrak{H}^0{}_+'$ respectively. Then by (2.13)

$$(2.18) \qquad [u,v] = \mp Q(u,v), \quad u \in \mathfrak{H}^0{}_\mp', \quad v \in \mathfrak{H}^0(L_1).$$

For our theory it is of fundamental importance that

$$(2.19) \qquad \mathfrak{H}^0(L_1) = \mathfrak{H}^0{}_+' + \mathfrak{H}^0{}_-',$$

where $\mathfrak{H}^0{}_+' + \mathfrak{H}^0{}_-'$ denotes the space of all functions

$$(2.20) \qquad u(x) = u_1(x) + u_2(x), \quad u_1(x) \in \mathfrak{H}^0{}_+', \quad u_2(x) \in \mathfrak{H}^0{}_-'.$$

Clearly the inclusion $\subset$ holds in (2.19).

In order to prove the inverse inclusion we first define two projection operators $P_0(x)$ and $N_0(x)$. Let the symmetric $m \times m$-matrix $P_0(x)$ project onto the subspace of all m-component vectors which is spanned by all eigenvectors corresponding to positive eigenvalues of $A(x)$. Accordingly let $N_0(x)$ project onto the subspace spanned by the eigenvectors corresponding to negative eigenvalues. Given now any $u(x) \in \mathfrak{H}^0(L_1)$ then by Lemma 1.1 the function $u_0(x) = P_0(x)u(x)$, $x \in \Gamma$, is a bounded measurable m-component vector function defined for $x \in \Gamma$. Further by Lemma 1.1 $A(x)u_0(x) = P(x)u_0(x) = \frac{1}{2}(A(x)u(x) + B(x)u(x))$ is Lipschitz continuous on Γ. Now we apply the following theorem which has been proved by the author in [5].

THEOREM 2.2. *Let the domain D, its boundary Γ, the operator L_1 in $\mathfrak{D}(L_1)$ and the matrix $A(x)$ satisfy all the assumptions mentioned above; especially let $\mathfrak{D}(L_1)$ be characterized by the conditions* a) *to* e). *Then, if $u_0(x)$ is any complex valued bounded measurable function defined on Γ and if*

$$v(x) = A(x)u_0(x)$$

is Lipschitz continuous on Γ, there exists an m-component vector function $u(x) \in \mathfrak{D}(L_1)$ such that $u(x) = u_0(x)$, $x \in \Gamma$.

Applying this theorem for our $u_0(x)$ we immediately obtain the existence of some $u_1(x) \in \mathfrak{D}(L_1)$ with $u_1(x) = u_0(x)$ on Γ. We define $u_2(x) = u(x) - u_1(x)$, $x \in D + \Gamma$. Clearly then $u_2(x) = (1 - P_0(x))u(x)$, $x \in \Gamma$, and therefore $N_0(x)u_2(x) = 0$, $x \in \Gamma$. Consequently (2.20) holds. This proves that

$$(2.21) \qquad \mathfrak{D}(L_1) \subset \mathfrak{H}^0{}_+' + \mathfrak{H}^0{}_-'$$

and therefore that equation (2.19) is true.

One should observe that application of the complicated Theorem 2.2 can be avoided if the matrix $A(x)$ is assumed to be of constant rank. This

follows because then, by a well known perturbation theorem of F. Rellich [16], the matrix $P_0(x)$ has continuous first derivatives. Hence the vector function $u_0(x)$ defined above also is Lipschitz continuous and its continuation into the interior of $D+\Gamma$ to a function $u_1(x) \in \mathfrak{D}(L_1)$ is trivial. In that case also one would be able to replace the domain $\mathfrak{D}(L_1)$ defined above by the much simpler domain $C^1(D+\Gamma)$ of all vector functions which are continuously differentiable on $D+\Gamma$, without disturbing any of the statements which follow in this paper.

Now for $u \in \mathfrak{H}^0{}_+'$, $v \in \mathfrak{H}^0(L_1)$ we obtain the estimate

$$|[u,v]| = |Q(u,v)| \leqq \| u \| \, \| v \|. \tag{2.22}$$

Therefore a vector function $u^0 \in \mathfrak{H}^0$ exists such that

$$[u,v] = (u^0, v), \qquad v \in \mathfrak{H}^0(L_1). \tag{2.23}$$

The correspondence $u \to u^0$, of course, is unique and linear and therefore defines a linear operator which is defined in $\mathfrak{H}^0{}_+'$. Analogously by use of (2.17) for $u \in \mathfrak{H}^0{}_-'$, $v \in \mathfrak{H}^0(L_1)$, the same inequality (2.22) can be shown and therefore again a $u^0 \in \mathfrak{H}^0$ exists such that (2.23) is true. This shows that for $u \in \mathfrak{H}^0{}_+' + \mathfrak{H}^0{}_-' = \mathfrak{H}^0(L_1)$ there exists always an element u^0 such that (2.23) holds. Simply decompose $u(x)$ in the way indicated by (2.33) and define $u^0 = u_1{}^0 + u_2{}^0$. We define

$$G^{0\prime} u = u^0, \qquad u \in \mathfrak{H}^0(L_1), \tag{2.24}$$

and then get a uniquely defined herimitian symmetric positive definite operator which satisfies

$$[u,v] = (G^{0\prime} u, v), \qquad u, v \in \mathfrak{H}^0(L_1). \tag{2.25}$$

Now by a well known theorem of K. O. Friedrichs every hermitian symmetric positive definite operator has a selfadjoint extension. Let

$$([u,v]) = (u,v) + [u,v] \tag{2.26}$$

and

$$([u]) = ([u,u])^{\frac{1}{2}}. \tag{2.27}$$

Let further $\mathfrak{P}^0$ be the completion of $\mathfrak{H}^0(L_1)$ with respect to the positive definite metric $([u])$. Then by Friedrichs one of these extensions of $G^{0\prime}$ is given by the restriction of $G^{0\prime *}$ in $\mathfrak{D}(G^{0\prime *})$ to $\mathfrak{P}^0 \cap \mathfrak{D}(G^{0\prime *})$. We chose this special extension to be our operator G^0 and then we proved not only Theorem 2.1, but also the following.

Corollary to Theorem 2.1: the operator G^0 in $\mathfrak{D}(G^0)$ can be chosen in such a way that

$$(2.28) \qquad \mathfrak{D}(G^0) \subset \mathfrak{P}^0, \quad [u,v] = (u, G^0 v), \quad u \in \mathfrak{P}^0, \quad v \in \mathfrak{D}(G^0).$$

Finally it is necessary to remark that for our theory it is very essential that already the operator $G^{0\prime}$ in $\mathfrak{D}(G^{0\prime}) = \mathfrak{H}^0(L_1)$ is essentially selfadjoint. This will be concluded in section 5 using a result about a special dissipative extension of L_0 in $\mathfrak{D}(L_0)$, characterized by local boundary conditions, which was first obtained by K. O. Friedrichs [7], R. S. Phillips and P. D. Lax [15].

3. The boundary space $\mathfrak{B}$. Let $\mathfrak{H}$ be the Hilbert space of all m-component vector functions $u(x)$ which are square integrable over D and let $\langle u, v\rangle$ denote the inner product as defined in (2.10). Let the operator L_0 in $\mathfrak{D}(L_0)$ be the restriction of L_1 in $\mathfrak{D}(L_1)$ to the space

$$(3.1) \qquad \mathfrak{D}(L_0) = \{u(x) \mid u \in \mathfrak{D}(L_1),\ u = 0 \text{ outside a compact subset of } D.\}.$$

Let $L_0{}^*$ in $\mathfrak{D}(L_0{}^*)$ be the (strict) adjoint of L_0 in $\mathfrak{D}(L_0)$. By definition $\mathfrak{D}(L_0{}^*)$ is the set of all $u \in \mathfrak{H}$ for which an element $u^* \in \mathfrak{H}$ exists such that $\langle u, L_0 v\rangle = \langle u^*, v\rangle$, $v \in \mathfrak{D}(L_0)$, and then by definition $L_0{}^* u = u^*$. Because of (2.2) the operator iL_0 certainly is hermitian symmetric and therefore we get $L_0{}^* \supset -L_0$; i.e., $\mathfrak{D}(L_0{}^*) \supset \mathfrak{D}(L_0)$, $L_0 u = -L_0 u$, $u \in \mathfrak{D}(L_0)$. More particularly we also get $L_0{}^* \supset -L_1$ as (2.13) shows. Since $L_0{}^*$ by definition is closed we get $L_0{}^* \supset -L_1{}^{**}$. It is a very essential fact for our theory that both of the operators of the last inclusion are equal:

$$(3.2) \qquad L_0{}^* = -L_1{}^{**}.$$

This was proved first by K. O. Friedrichs [6]. Friedrichs calls the solutions $u \in \mathfrak{D}(L_0{}^*)$ of $-L_0{}^* u = f$ weak solutions, the solutions $u \in \mathfrak{D}(L_1{}^{**})$ of $L_1{}^{**} u = f$ strong solutions of the differential equation $L_1 u = f$. Using a certain type of integral operators, the so called mollifiers, he proves in his paper that every weak solution is also a strong solution, i.e., that $L_0{}^* \subset -L_1{}^{**}$. One can easily check that the assumptions for his theorem here are satisfied.

We now introduce the boundary space $\mathfrak{B}$ by

$$(3.3) \qquad \mathfrak{B} = \mathfrak{H}^0 \ominus \mathfrak{H}^0(L_0{}^{**}),$$

where the orthogonal complement is taken with respect to the inner product (u, v) defined in (2.6). Here $\mathfrak{H}^0(L_0{}^{**})$ denotes the closed subspace of $\mathfrak{H}^0$ which as a set of elements is equal to $\mathfrak{D}(L_0{}^{**})$. Using (3.8) it can easily be seen that the space $\mathfrak{B}$ introduced by this definition is isomorphic to the boundary space $\mathfrak{B}$ introduced by Calkin [2]. By (2.13) we obtain the relation

$$(3.4) \qquad Q(u,v) = 0, \qquad u \in \mathfrak{H}^0, \qquad v \in \mathfrak{H}^0(L_0{}^{**}).$$

Since $\mathfrak{H}^0(L_0^{**})$ and $\mathfrak{B}$ by definition are closed under the norm $\|u\|$, every element $\mathfrak{H}^0$ can be decomposed uniquely:

$$(3.5) \qquad u = \phi + u_0, \qquad \phi \in \mathfrak{B}, \qquad u_0 \in \mathfrak{H}^0(L_0^{**}).$$

Therefore if $u, v \in \mathfrak{H}^0$ and $u = \phi + u_0$, $v = \psi + v_0$ are the corresponding decompositions then by (3.3):

$$(3.6) \qquad Q(u,v) = Q(\phi,\psi).$$

Hence the value of the form $Q(u,v)$, $u, v \in \mathfrak{H}^0$ is already determined by the orthogonal components of u and v in $\mathfrak{B}$. Therefore it sufficies to consider $Q(u,v)$ for $u, v \in \mathfrak{B}$.

The notation "boundary space" is justified by the following: If

$$u = \phi + u_0, \quad u_0 \in \mathfrak{H}^0(L_0^{**}), \quad \phi \in \mathfrak{B}, \quad v = \phi + v_0, \quad v_0 \in \mathfrak{H}^0(L_0^{**}),$$

then

$$(3.7) \qquad u - v = u_0 - v_0 \in \mathfrak{D}(L_0^{**}).$$

The elements of $\mathfrak{D}(L_0^{**})$ vanish at the boundary in some generalized sense. Therefore the function ϕ determines the boundary values of u in some generalized manner. One may say that u and v have the same boundary values in this sense.

According to the theory developed by R. S. Phillips in [14] any closed operator M which is an extension of L_0^{**} and a contradiction of $-L_0^*$ corresponds to a closed subspace

$$\mathfrak{B}(M) = \mathfrak{H}^0(M) \ominus \mathfrak{H}^0(L_0^{**}) \text{ of } \mathfrak{B}.$$

This correspondence is one to one. If M_1, M_2 are two such operators then $M_1 \subset M_2$ implies $\mathfrak{B}(M_1) \subset \mathfrak{B}(M_2)$. An operator M in $\mathfrak{D}(M)$ is called dissipative if $\langle u, Mu\rangle + \langle Mu, u\rangle \leqq 0$, $u \in \mathfrak{D}(M)$. It is called maximal dissipative if it is dissipative and if it does not have any proper extension which is also dissipative. M is dissipative if and only if $\mathfrak{B}(M)$ is negative with respect to the form $Q(u,u)$:

$$(3.8) \qquad Q(u,u) \leqq 0, \qquad u \in \mathfrak{B}(M).$$

M is maximal dissipative if and only if $\mathfrak{B}(M)$ is maximal negative with respect to $Q(u,u)$, i.e., if every negative extension of $\mathfrak{B}(M)$ coincides with $\mathfrak{B}(M)$. iM is selfadjoint if and only if $\mathfrak{B}(M)$ is a nullspace of $Q(u,u)$ which is maximal negative and maximal positive:

$$Q(u,u) = 0, \qquad u \in \mathfrak{B}(M);$$

there are no proper extensions of $\mathfrak{B}(M)$ which are entirely positive or entirely negative.

In order to get more information about the space $\mathfrak{B}$ we remember that by definition $\mathfrak{B}$ is the space of all $\phi \in \mathfrak{H}^0$ such that

$$(\phi, u) = \langle\phi, u\rangle + \langle L_1^{**}\phi, L_0^{**}u\rangle = 0, \qquad u \in \mathfrak{D}(L_0^{**}). \tag{3.9}$$

Hence

$$L_1^{**}\phi \in \mathfrak{D}(L_0^{*}) = \mathfrak{D}(L_1^{**})$$

and

$$L_1^{**}(L_1^{**}\phi) = \phi. \tag{3.10}$$

Here again (3.2) was used. Consequently $\mathfrak{B}$ is the space of all solutions of

$$(L_1^{**})^2\phi = \phi. \tag{3.11}$$

Since $\phi \in \mathfrak{B}$ implies $\phi \in \mathfrak{D}(L_1^{**})$, we obtain by applying L_1^{**} to (3.11) that

$$L_1^{**}\phi \in \mathfrak{B} \text{ for every } \phi \in \mathfrak{B}. \tag{3.12}$$

Therefore the operator L_1^{**} transforms the space $\mathfrak{B}$ into itself. We denote the restriction of L_1^{**} to $\mathfrak{B}$ by L; then (3.11) implies

$$L^2 = 1. \tag{3.13}$$

On the other hand by (2.13) and (3.11) for $u, v \in \mathfrak{B}$:

$$\begin{aligned}(Lu, v) &= \langle Lu, v\rangle + \langle L^2u, Lv\rangle = \langle Lu, v\rangle + \langle u, Lv\rangle \\ &= Q(u, v) = (u, Lv).\end{aligned} \tag{3.14}$$

Further, if $u, v \in \mathfrak{B}$, then

$$(Lu, Lv) = \langle Lu, Lv\rangle + \langle L^2u, L^2v\rangle = \langle Lu, Lv\rangle + \langle u, v\rangle = (u, v). \tag{3.15}$$

Hence L is a hermitian symmetric and unitary operator of the space $\mathfrak{B}$ into itself. In other words, L is a symmetry. Also we note that

$$Q(u, v) = (Lu, v), \qquad u, v \in \mathfrak{B}. \tag{3.16}$$

Consequently L has eigenvalues at $\lambda = +1$ and $\lambda = -1$ at most. (For special operators L_1 it may happen that one or both of the corresponding eigenspaces do not contain any element different from zero.) The identity

$$u = \tfrac{1}{2}(1 + L)u + \tfrac{1}{2}(1 - L)u \tag{3.17}$$

gives an eigenfunction expansion of the arbitrary element $u \in \mathfrak{B}$. The operators

$$\tfrac{1}{2}(1 + L), \qquad \tfrac{1}{2}(1 - L) \tag{3.18}$$

are the projections onto the eigenspaces belonging to $\lambda=+1$ and $\lambda=-1$ respectively. Further we note that

$$(3.19)\quad Q(u,v)=\tfrac{1}{4}((1+L)u,(1+L)v)-\tfrac{1}{4}((1-L)u,(1-L)v)$$

for every $u,v\in\mathfrak{B}$. Using these facts Phillips also proves the following

THEOREM 3.1. *A negative subspace $\mathfrak{B}(M)$ of $\mathfrak{B}$ is maximal negative with respect to $Q(u,u)$ if and only if*

$$(3.20)\qquad (1-L)\mathfrak{B}(M)=(1-L)\mathfrak{B}.$$

We repeat the proof: First of all a maximal negative subspace must be closed under the norm of $\mathfrak{B}$, because otherwise the closure would be a proper negative extension. Since L has its spectrum only at $\lambda=\pm 1$, the space $(1-L)\mathfrak{B}$ is closed and the space $(1-L)\mathfrak{B}(M)$ must be closed also, because $u^n\in\mathfrak{B}(M)$, $(1-L)u^n\to v$ implies

$$\begin{aligned}0&\geqq 4Q(u^n-u^m,u^n-u^m)\\&=\|(1+L)(u^n-u^m)\|^2-\|(1-L)(u^n-u^m)\|^2\end{aligned}$$

and therefore

$$(3.21)\quad \|(1+L)(u^n-u^m)\|^2\leqq\|(1-L)(u^n-u^m)\|^2\to 0,\quad n,m\to\infty.$$

Hence

$$(3.22)\qquad u_n=\tfrac{1}{2}(1+L)u_n+\tfrac{1}{2}(1-L)u_n$$

converges too. Let $u=\lim_{n\to\infty}u_n$; then u is in the closure of $\mathfrak{B}(M)$ and therefore in $\mathfrak{B}(M)$. Hence $v=(1-L)u$ for an element $u\in\mathfrak{B}(M)$. Therefore if $(1-L)\mathfrak{B}(M)\subset(1-L)\mathfrak{B}$ is a proper inclusion, the complement $(1-L)\mathfrak{B}\ominus(1-L)\mathfrak{B}(M)$ will contain a vector $\phi\neq 0$. But then

$$Q(\phi+u,\phi+u)=Q(u,u)-\tfrac{1}{4}((1-L)\phi,(1-L)\phi)\leqq 0,\quad u\in\mathfrak{B}(M).$$

Hence a proper negative extension exists and $\mathfrak{B}(M)$ can not be a maximal negative.

Conversely, if (3.20) holds, then $\mathfrak{B}(M)$ is maximal negative. For let $\mathfrak{B}(M^0)$ be any negative extension of $\mathfrak{B}(M)$ and let $\phi\in\mathfrak{B}(M^0)$. Then there exists a $\psi\in\mathfrak{B}(M)$ such that $(1-L)(\phi-\psi)=0$. But $\phi-\psi\in\mathfrak{B}(M^0)$; hence

$$\begin{aligned}0\geqq 4Q(\phi-\psi,\phi-\psi)&=\|(1+L)(\phi-\psi)\|^2-\|(1-L)(\phi-\psi)\|^2\\&=\|(1+L)(\phi-\psi)\|^2.\end{aligned}$$

Consequently $(1+L)(\phi-\psi)=0$, $\phi=\psi$, $\mathfrak{B}(M^0)=\mathfrak{B}(M)$. Therefore Theorem 3.1 is proved. Especially we also get the following

COROLLARY. *The closure of a negative subspace* $\mathfrak{B}(M)$ *is maximal negative if and only if* $(1-L)\mathfrak{B}(M)$ *is dense in* $(1-L)\mathfrak{B}$.

4. The boundary space $\mathfrak{E}$ and the space $\mathfrak{P}$. When we consider the form $[u,v]$ and the form $Q(u,v)$ in their special expressions as boundary integrals then it seems to be more natural to consider both forms in another boundary space, which also would have more right to be called boundary space, namely in some space which consists of functions defined only on the boundary Γ.

Let $A(x)$ and $B(x)$ be the $m\times m$-matrices defined in (2.3) and (2.4). Let $\mathfrak{E}''$ be the space of all (classes of equivalent) bounded measurable m-component vector functions $u(x)$ defined for $x\in\Gamma$ only for which $A(x)u(x)$ is Lipschitz continuous on Γ, and let $\mathfrak{Z}''$ be the subspace of $\mathfrak{E}''$ consisting of all those $u\in\mathfrak{E}''$ for which $A(x)u(x)=0$ on Γ. Then by (2.5) and (2.13) the forms $[u,v]$ and $Q(u,v)$ can be defined for $u,v\in\mathfrak{E}''$ and we get

$$(4.1)\qquad [u,v]=0,\qquad Q(u,v)=0,\qquad \text{if } u\in\mathfrak{E}'', v\in\mathfrak{Z}''.$$

It is easy to see that $\mathfrak{Z}''$ is just the nullspace of the form $[u,u]$ in $\mathfrak{E}''$. Therefore, if we define

$$(4.2)\qquad \mathfrak{E}'=\mathfrak{E}''/\mathfrak{Z}''$$

then the forms $[u,v]$ and $Q(u,v)$ induce two corresponding forms in $\mathfrak{E}'$ and the value of $[u,v]$ and $Q(u,v)$ for $u,v\in\mathfrak{E}''$ already is determined by the residue classes of u and v with respect to this factorization. Clearly the form $[u,u]$ is positive definite in $\mathfrak{E}'$. We now complete $\mathfrak{E}'$ with respect to the norm

$$(4.3)\qquad [[u]]=\{[u,u]\}^{\frac{1}{2}}$$

and then call the completed space $\mathfrak{E}$.

We would like to compare this new boundary space with the boundary space $\mathfrak{B}$, defined in section 3. For this purpose let L_0' in $\mathfrak{D}(L_0')$ be the restriction of L_1 in $\mathfrak{D}(L_1)$ to the space $\mathfrak{D}(L_0')$ of all $u\in\mathfrak{D}(L_1)$ which, restricted to the boundary, belong to the space $\mathfrak{Z}''$. Clearly $L_0\subset L_0'$ and (2.13) yields $L_0'\subset -L_1^*=L_0^{**}$. On the other hand

$$(4.4)\qquad \mathfrak{E}'\cong\mathfrak{D}(L_1)/\mathfrak{D}(L_0').$$

We can introduce the metric $\|u\|$ for elements $u\in\mathfrak{E}'$ by defining $\|u\|$ to be equal to the greatest lower bound of the $\|\ \|$-norms of all elements of $\mathfrak{D}(L_1)$ which belong to the residue class corresponding to u. By setting

$$(u,v)=\tfrac{1}{4}\{\|u+v\|^2-\|u-v\|^2-i\|u+iv\|^2+i\|u-iv\|^2\}$$

we also get the form (u,v) defined for $u,v\in\mathfrak{E}'$. We prove

LEMMA 4.1.

$$(4.5) \qquad (u,u) > 0, \qquad u \neq 0, \qquad u \in \mathfrak{E}'.$$

In other words, (u,u) is positive definite also in $\mathfrak{E}'$.

Proof. Let $(u,u) = \| u \|^2 = 0$ for some nonvanishing $u \in \mathfrak{E}'$. Then there exists a sequence $u^n = u^0 - v^n$, $n = 1, 2, \cdots$, such that $u^0 \in \mathfrak{D}(L_1)$, $u^0 \notin \mathfrak{D}(L_0')$, $v^n \in \mathfrak{D}(L_1')$ and $\lim_{n\to\infty} \| u^n \| = 0$. Since $Q(u,v) = 0$ for $u \in \mathfrak{D}(L_0')$, $v \in \mathfrak{D}(L_1)$, and because the form is continuous with respect to the metric $\| u \|$ this means that $Q(u^0, v) = \lim_{n\to\infty} Q(u^n, v) = 0$ for every $v \in \mathfrak{D}(L_1)$. But by Theorem 2.2 there exists an element $v^0 \in \mathfrak{D}(L_1)$ with $v^0(x) = A(x)u^0(x)$, $x \in \Gamma$. The above limit for $v = v^0$ and (2.13) furnish

$$Q(u^0, v^0) = \int_\Gamma | A(x) u^0(x) |^2 \, d\sigma = 0$$

or $A(x)u^0(x) = 0$ a.e. on Γ. This means that $u^0(x) \in \mathfrak{D}(L_0')$ which is a contradiction. Therefore Lemma 4.1 is proved.

The next conclusion is

LEMMA 4.2.

$$(4.6) \qquad \mathfrak{E}' \cong \mathfrak{D}(L_1)/\mathfrak{D}(L_0') \cong \mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**}),$$

where $\mathfrak{H}^0(L_1)$ is defined as in Section 2, and where $\mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**})$ means the set of the $\mathfrak{B}$-components of elements of $\mathfrak{H}^0(L_1)$ according to the decompostition (3.5). The isomorphy includes the norm $\| u \|$ and the form $Q(u,v)$.

Proof. Since $\mathfrak{D}(L_0') \cong \mathfrak{H}^0(L_0')$ and $\mathfrak{H}^0(L_0') \subset \mathfrak{H}^0(L_0^{**})$ all elements of a certain residue class of $\mathfrak{D}(L_1)/\mathfrak{D}(L_0')$ have the same projection onto the space $\mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**})$. We establish the isomorphy between the above two spaces by assigning to each residue class $u \in \mathfrak{D}(L_1)/\mathfrak{D}(L_0')$ the common projection of its elements onto the space $\mathfrak{H}^0(L_1') \ominus \mathfrak{H}^0(L_0^{**})$. Clearly this defines a homomorphism of $\mathfrak{D}(L_1)/\mathfrak{D}(L_0')$ onto $\mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**})$ which preserves the vector operations as well as the norm $\| u \|$ and the form $Q(u,v)$. Hence we only have to show that the correspondence is one to one. Now, if u and v are different residue classes of $\mathfrak{D}(L_1)/\mathfrak{D}(L_0')$, then by Lemma 4.1 we get $\| u - v \| \neq 0$. But if u and v correspond to the same element of $\mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**})$, then $u^0 - v^0 \in \mathfrak{D}(L_0^{**})$ for $u^0 \in u$, $v^0 \in v$, and thus we will be able to find a sequence $w^n \in \mathfrak{D}(L_0')$ with $\lim_{n\to\infty} \| u^0 - v^0 - w^n \| = 0$. But this means that $\| u - v \| = 0$ and therefore we get a contradiction.

For convenience we introduce the notation

$$(4.7) \qquad \mathfrak{B}' = \mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**}).$$

Then Lemma 4.2 can be expressed in the form

$$(4.8) \qquad \mathfrak{E}' \cong \mathfrak{B}'.$$

Now we can complete $\mathfrak{E}'$ also with respect to the norm $\| u \|$. Since this completion obviously is isomorphic to the space $\mathfrak{H}^0 \ominus \mathfrak{H}^0(L_0^{**}) = \mathfrak{B}$ we can state

LEMMA 4.3. *The completion of $\mathfrak{E}'$ with respect to the metric $\| u \|$ is isomorphic to the boundary space $\mathfrak{B}$, the isomorphy including the metric $\| u \|$ and the form $Q(u,v)$.*

As a special fact we note that the isomorphy (4.8) introduces the form $[u,v]$ also for the dense subspace $\mathfrak{B}'$ of the boundary space $\mathfrak{B}$.

We find it convenient to introduce as a third positive definite inner product the form

$$(4.9) \qquad ([u,v]) = (u,v) + [u,v], \qquad u, v \in \mathfrak{E}' \cong \mathfrak{B}'.$$

We can complete the space $\mathfrak{E}' \cong \mathfrak{B}'$ also with respect to the metric

$$(4.10) \qquad ([u]) = \{([u,u])\}^{\frac{1}{2}}.$$

We denote this completion by $\mathfrak{P}$ and its dense subspace corresponding to $\mathfrak{E}' \cong \mathfrak{B}'$ by $\mathfrak{P}'$. When we consider the space $\mathfrak{P}^0$ defined in Section 2, then we find between $\mathfrak{P}$, $\mathfrak{P}'$ and $\mathfrak{P}^0$ the following relation. Clearly $\mathfrak{P}^0 \supset \mathfrak{H}^0(L_0')$ and since $[u,u] = 0$, $u \in \mathfrak{H}^0(L_0')$, we get $([u,u]) = (u,u)$, $u \in \mathfrak{H}^0(L_0')$. This shows that there is a closed subspace $\mathfrak{P}(L_0^{**})$ of $\mathfrak{P}^0$ which corresponds elementwise to $\mathfrak{H}^0(L_0^{**})$. Now we state

LEMMA 4.4.

$$(4.11) \qquad \mathfrak{P}' \cong \mathfrak{P}^0(L_1) \ominus \mathfrak{P}^0(L_0^{**}), \qquad \mathfrak{P} \cong \mathfrak{P}^0 \ominus \mathfrak{P}^0(L_0^{**}),$$

where the orthogonal complement is taken with respect to the inner product $([u,v])$.

Proof. We only note that $([u,v]) = (u,v)$ for $u \in \mathfrak{P}^0$, $v \in \mathfrak{P}^0(L_0^{**})$. Therefore

$$\mathfrak{P}^0(L_1) \ominus \mathfrak{P}^0(L_0^{**}) \cong \mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**}) \cong \mathfrak{E}' \cong \mathfrak{B}' \cong \mathfrak{P}',$$

where the first orthogonal complement is taken with respect to the inner product $([u,v])$ but the second with respect to the inner product (u,v). Consequently the first formula (4.11) is proved and the second formula now

immediately follows from the fact that $\mathfrak{P}$ and $\mathfrak{P}^0$ are defined as completions of $\mathfrak{P}'$ and $\mathfrak{P}^0(L_1) \cong \mathfrak{H}^0(L_1)$ respectively. Especially we get the following

COROLLARY. *Orthogonality of $u, v \in \mathfrak{P}^0$ with respect to (u, v) and with respect to $([u, v])$ means the same if at least one of the elements u, v is contained in $\mathfrak{P}^0(L_0^{**}) \cong \mathfrak{H}^0(L_0^{**})$.*

5. The operator G in $\mathfrak{D}(G)$. Next we consider the form $Q(u, v)$ in the space $\mathfrak{E}$. We refer to the definition of the spaces $\mathfrak{H}^0{}_+'$ and $\mathfrak{H}^0{}_-'$ in Section 2. $\mathfrak{H}^0{}_+'$, for instance, was the space of all $u \in \mathfrak{H}^0(L_1)$ with

$$N(x)u(x) = \tfrac{1}{2}(B(x) - A(x))u(x) = 0 \text{ on } \Gamma, \tag{5.1}$$

and $\mathfrak{H}^0{}_-'$, the corresponding space with $N(x)$ replaced by $P(x)$. We define

$$\mathfrak{E}_\pm' = \mathfrak{H}^0{}_\pm' \, \mathfrak{H}^0(L_0^1). \tag{5.2}$$

Clearly $\mathfrak{E}' = \mathfrak{E}_+' \oplus \mathfrak{E}_-'$ under $[u, v]$. The relation (2.18) yields

$$[u, v] = \pm Q(u, v), \qquad u \in \mathfrak{E}_\pm', \qquad v \in \mathfrak{E}'. \tag{5.3}$$

If we denote the closures of $\mathfrak{E}_+'$ and $\mathfrak{E}_-'$ with respect to $[[u]]$ by $\mathfrak{E}_+$ and $\mathfrak{E}_-$ respectively, then

$$\mathfrak{E} = \mathfrak{E}_+ \oplus \mathfrak{E}_- \text{ under } [u, v]. \tag{5.4}$$

By (5.3) the form $Q(u, v)$ can be extended continuously to all $u \in \mathfrak{E}_+$, $v \in \mathfrak{E}$, and analogously to all $u \in \mathfrak{E}_-$, $v \in \mathfrak{E}$. Since $\mathfrak{E}_+$ and $\mathfrak{E}_-$ are orthogonal complements of each other with respect to $[u, v]$, we can define a bounded operator Q in $\mathfrak{E}$ by

$$Qu = \pm u, \qquad u \in \mathfrak{E}_\pm. \tag{5.5}$$

We then get

LEMMA 5.1. *There exists a bounded selfadjoint unitary operator Q defined in all of $\mathfrak{E}$ and having its spectrum only at $\lambda = \pm 1$ at the most, satisfies the relation*

$$Q(u, v) = [Qu, v], \qquad u, v \in \mathfrak{E}. \tag{5.6}$$

We now consider the operator G^0 in $\mathfrak{D}(G^0)$ defined in Theorem 2.1. First we again observe that $\mathfrak{H}^0(L_0')$ and also its closure $\mathfrak{H}^0(L_0^{**})$ are null-spaces of G^0. This follows because for $\phi = \mathfrak{H}^0(L_0')$, $v \in \mathfrak{H}^0(L_1)$, Theorem 2.1 and formula (2.13) imply $(G^0\phi, v) = [\phi, v] = 0$.

Consequently the selfadjoint operator G^0 in $\mathfrak{D}(G^0)$ transforms the orthogonal complement $\mathfrak{B}$ of $\mathfrak{H}^0(L_0^{**})$ into itself. Hence the restriction of G^0 to

$\mathfrak{D}(G) = \mathfrak{B} \cap \mathfrak{D}(G^0)$ considered as a transformation of $\mathfrak{B}$ into itself is a selfadjoint operator. We call this operator G in $\mathfrak{D}(G)$.

LEMMA 5.4.

$$\mathfrak{B}' \subset \mathfrak{D}(G) \subset \mathfrak{P}. \tag{5.7}$$

Proof. By definition $\mathfrak{B}' = \mathfrak{H}^0(L_1) \ominus \mathfrak{H}^0(L_0^{**})$, $\mathfrak{D}(G) = \mathfrak{D}(G^0) - \mathfrak{H}^0(L_0^{**})$, and $\mathfrak{P} = \mathfrak{P}^0 \ominus \mathfrak{H}^0(L_0^{**})$. But by Theorem 2.1 and its corollary, $\mathfrak{H}^0(L_1) \subset \mathfrak{D}(G^0) \subset \mathfrak{P}^0$. This proves the lemma.

LEMMA 5.5.

$$[u, v] = (u, Gv), \qquad u \in \mathfrak{P}, \qquad v \in \mathfrak{D}(G). \tag{5.8}$$

The proof is an obvious consequence of (2.28) and (5.7).

LEMMA 5.6. *G in $\mathfrak{D}(G) \subset \mathfrak{B}$ has a densely defined inverse.*

Proof. Let $G\phi = 0$, $\phi \in \mathfrak{D}(G)$. Then by (5.8), $[\phi, v] = (G\phi, v) = 0$, $v \in \mathfrak{P}$. But $\mathfrak{P}$ is dense in $\mathfrak{E}$, hence $\phi = 0$. Let G' in $\mathfrak{D}(G')$ be the restriction of G in $\mathfrak{D}(G)$ to $\mathfrak{D}(G') = \mathfrak{B}'$.

THEOREM 5.1. *The operator G' in $\mathfrak{D}(G')$ is essentially selfadjoint.*

Proof. Let ϕ be an element of $\mathfrak{B}$ satisfying

$$(\phi, (G' + 1)u) = 0, \qquad u \in \mathfrak{D}(G'). \tag{5.9}$$

Then

$$(\phi, (G^{0\prime} + 1)u) = 0, \qquad u \in \mathfrak{D}(G^{0\prime}), \tag{5.9a}$$

since $\mathfrak{D}(G^{0\prime}) \subset \mathfrak{D}(G') + \mathfrak{H}^0(L_0^{**})$ and $\mathfrak{H}^0(L_0^{**}) \perp \mathfrak{D}(G')$ under the inner product (u, v). Since $\mathfrak{D}(G^{0\prime}) = \mathfrak{H}^0{}_+' + \mathfrak{H}^0{}_-'$, (5.9a) holds for $u \in \mathfrak{H}^0{}_+'$ and $u \in \mathfrak{H}^0{}_-'$. For $u \in \mathfrak{H}^0{}_+'$, by definition $(\phi, G'u) = Q(\phi, u) = \langle L\phi, u\rangle + \langle \phi, Lu\rangle$ and on the other hand $(\phi, u) = \langle \phi, u\rangle + \langle L\phi, L_1 u\rangle$. Hence (5.9a), for $u \in \mathfrak{H}^0{}_+'$, is equivalent to

$$\langle L\phi, u\rangle + \langle \phi, L_1 u\rangle + \langle \phi, u\rangle + \langle L\phi, L_1 u\rangle = 0$$

or

$$\langle (1 + L)\phi, (1 + L_1)u\rangle = 0, \qquad u \in \mathfrak{H}^0{}_+'.$$

Let L_+ in $\mathfrak{D}(L_+)$ be the restriction of L_1 in $\mathfrak{D}(L_1)$ to the space $\mathfrak{D}(L_+) = \mathfrak{H}^0{}_+' \subset \mathfrak{D}(L_1)$. Then it follows that

$$\langle (1 + L)\phi, (1 + L_+)u\rangle = 0, \qquad u \in \mathfrak{D}(L_+). \tag{5.10}$$

Hence

$$(1 + L)\phi \in \mathfrak{D}(L_+^*), \ (1 + L_+^*)(1 + L)\phi = 0. \tag{5.11}$$

Now

$$L_+^* = -L_-^{**}, \tag{5.12}$$

where L_- in $\mathfrak{D}(L_-)$ is the restriction of L_1 in $\mathfrak{D}(L_1)$ to $\mathfrak{D}(L_-) = \mathfrak{H}^0{}_-'$. This follows from a theorem which first was proved by K. O. Friedrichs [6] and then was extended by R. S. Phillips and P. D. Lax [15] to the generality required here.* Hence

$$(1 - L_-^{**})(1 + L)\phi = 0. \tag{5.13}$$

But for $u \in \mathfrak{D}(L_-)$ we get

$$\begin{aligned} \langle\langle(1 - L_-u\rangle\rangle & \\ &= \langle u, u\rangle + \langle L_-u, L_-u\rangle - \langle L_-u, u\rangle - \langle u, L_-u\rangle \\ &= (u, u) - Q(u, u) \geqq (u, u) \geqq \langle u, u\rangle \end{aligned}$$

since $Q(u, u) \leqq 0$, $u \in \mathfrak{D}(L_-) = \mathfrak{H}^0{}_-'$. By closing the space $\mathfrak{D}(L_-)$ we get

$$\langle\langle(1 - L_-^{**})u\rangle\rangle \geqq \langle\langle u\rangle\rangle, \qquad u \in \mathfrak{D}(L_-^{**}). \tag{5.14}$$

Here we substitute $u = (1 + L)\phi$. Then $(1 - L_-^{**})u = 0$. Hence

$$(1 + L)\phi = 0. \tag{5.15}$$

In the analogous manner we conclude that $(1 - L)\phi = 0$, using (5.9a) for $u \in \mathfrak{D}(L_-) = \mathfrak{H}^0{}_-'$. Hence

$$\phi = \tfrac{1}{2}(1 + L)\phi + \tfrac{1}{2}(1 - L)\phi = 0. \tag{5.16}$$

This shows that the inverse $(1 + G')^{-1}$ is densely defined. Since on the other hand G' is positive we get that $(1 + G')^{-1}$ is bounded. Hence G' in $\mathfrak{D}(G')$ is essentially selfadjoint and the theorem is proved.

Next we state

THEOREM 5.2. *The operator G in $\mathfrak{D}(G)$ satisfies the relation*

$$G^{-1} = LGL. \tag{5.17}$$

i. e.,

$$L\mathfrak{D}(G^{-1}) = \mathfrak{D}(G)$$

$$G^{-1}u = LGLu, \qquad u \in \mathfrak{D}(G^{-1}). \tag{5.18}$$

Proof. First we note that

$$Gu = \begin{cases} Lu, & u \in \mathfrak{E}_+' \\ -Lu, & u \in \mathfrak{E}_-'. \end{cases} \tag{5.19}$$

---

* This paper still will not cover the case where the matrix $A(x)$ defined in (2.3) changes rank on Γ. A paper of the author concerned with this case is in preparation.

This follows because for $u \in \mathfrak{E}_{\pm}'$, $v \in \mathfrak{E}'$, the relation

$$(5.20) \qquad (Gu, v) = [u, v] + \pm Q(u, v) = (\pm Lu, v)$$

holds, Now (5.19) and the relation $L^2 = 1$ yield

$$(5.21) \qquad G'^{-1} = \begin{cases} Lu, & Lu \in \mathfrak{E}_+' \\ -Lu, & Lu \in \mathfrak{E}_-'. \end{cases}$$

Replacing u by Lu in (5.21) we obtain

$$(5.22) \qquad G'^{-1}Lu = \begin{cases} u, & u \in \mathfrak{E}_+' \\ -u, & u \in \mathfrak{E}_-'. \end{cases}$$

Hence

$$(5.23) \qquad LG'^{-1}Lu = \begin{cases} Lu, & u \in \mathfrak{E}_+' \\ -Lu, & u \in \mathfrak{E}^{-\prime}. \end{cases}$$

and therefore

$$(5.24) \qquad \begin{aligned} &Lu \in \mathfrak{D}(G^{-1}), \quad u \in \mathfrak{D}(G'), \\ &LG'Lu = G'u, \quad u \in \mathfrak{D}(G'). \end{aligned}$$

Now, given any $u \in \mathfrak{D}(G)$ there exists a sequence $u^n \in \mathfrak{D}(G')$ with $u^n \to u$, $Gu^n \to Gu$, because by Theorem 5.1 the operator G' in $\mathfrak{D}(G')$ is essentially selfadjoint. Hence $LG^{-1}Lu^n = Gu^n \to Gu$. Since L is bounded, Lu^n converges to Lu. Also $G^{-1}Lu^n \to G^{-1}Lu$. Consequently $Lu \in \mathfrak{D}(G^{-1})$ and $LG^{-1}Lu = Gu$, $u \in \mathfrak{D}(G)$. Hence $LG^{-1}L = G$ and $G^{-1} = LGL$ which proves Theorem 5.2.

Let E_λ be the spectral resolution of G in $\mathfrak{D}(G)$. Assume E_λ to be continuous from the right. Then set

$$(5.25) \qquad F_\lambda = E_\lambda - E_{\lambda^{-1}-0}, \qquad \lambda > 1.$$

Obviously F_λ is a projection for every $\lambda > 1$ and this projection is orthogonal under all three of theinner products (u, v), $[u, v]$ and $([u, v])$.

THEOREM 5.3. *F_λ commutes with L:*

$$(5.26) \qquad F_\lambda L = LF_\lambda, \qquad \lambda > 1.$$

Proof. The relation $G^{-1} = LGL$ yields $f(G^{-1}) = Lf(G)L$, i.e.,

$$\int_0^\infty f(\lambda^{-1})\,dE_\lambda = L\Big(\int_0^\infty f(\lambda)\,dE_\lambda\Big)L$$

first for analytic functions, then, by passing to the limit, also for any arbitrary piecewise continuous function $f(\lambda)$. Now set

$$(5.27) \qquad f_\mu(\lambda) = \begin{cases} 1, & \lambda \leqq \mu \\ 0, & \text{elsewhere} \end{cases}$$

then

$$f_\mu(\lambda^{-1}) = \begin{cases} 1, & \lambda \geqq \mu^{-1} \\ 0, & \text{elsewhere.} \end{cases} \tag{5.28}$$

Hence

$$f_\mu(G^{-1}) = \int_0^\infty f_\mu(\lambda^{-1})\, dE_\lambda = \int_{\mu^{-1}-0}^\infty dE_\lambda = 1 - E_{\mu^{-1}-0}. \tag{5.29}$$

Consequently $1 - E_{\mu^{-1}-0} = LE_\mu L$ or $(1 - E_{\mu^{-1}-0})L = LE_\mu$, $0 < \mu < \infty$. Finally we observe that for $\lambda > 1: F_\lambda = E_\lambda(1 - E_{\lambda^{-1}-0})$. Hence

$$F_\lambda L = E_\lambda L E_\lambda = L(1 - E_{\lambda^{-1}-0})E_\lambda = LF_\lambda.$$

This proves the theorem.

6. Boundary conditions in $\mathfrak{E}$. Let P_0 and N_0 be the projections of $\mathfrak{E}$ onto the closures $\mathfrak{E}$ onto the closures $\mathfrak{E}_+$ and $\mathfrak{E}_-$of $\mathfrak{E}_+'$ and $\mathfrak{E}_-'$ respectively. Obviously these projections are locally given by

$$P_0 u = P_0(x)u(x), \qquad N_0 u = N_0(x)u(x), \tag{6.1}$$

where $P_0(x)$ and $N_0(x)$ are the matrices defined in Section 2. Now we state

THEOREM 6.1.

$$\begin{aligned} P_0 u &= \tfrac{1}{2}(1 + G^{-1}L)u, \qquad u \in \mathfrak{D}(G), \\ N_0 u &= \tfrac{1}{2}(1 - G^{-1}L)u, \qquad u \in \mathfrak{D}(G). \end{aligned} \tag{6.2}$$

This for $u \in \mathfrak{D}(G')$ simply follows from (5.22). For $u \in \mathfrak{D}(G)$ we simply obtain it by choosing a sequence $u^n \in \mathfrak{D}(G')$ with $u^n \to u$, $Gu^n \to Gu$ under $\| u \|$ and passing to the limit $n \to \infty$. Now let $\mathfrak{P}_+$ and $\mathfrak{P}_-$ be the closure of $\mathfrak{E}_+'$ and $\mathfrak{E}_-'$ under the norm of $\mathfrak{P}$.

THEOREM 6.2.

$$\mathfrak{D}(G) = \mathfrak{P}_+ + \mathfrak{P}_-. \tag{6.3}$$

Proof. Let $u^n \in \mathfrak{E}_+'$, $u^n \to u$ under the norm of $\mathfrak{P}$. By (5.21) $u^n \in \mathfrak{E}_+'$ implies $Gu^n = Lu^n \to Lu$. Hence Gu^n converges also and therefore $u \in \mathfrak{D}(G)$. Hence $\mathfrak{P}_+ \subset \mathfrak{D}(G)$ and analogously $\mathfrak{P}_- \subset \mathfrak{D}(G)$.

Finally let $u \in \mathfrak{D}(G)$; then $u = P_0 u + N_0 u$ and

$$\begin{aligned} P_0 u &= \tfrac{1}{2}(1 + G^{-1}L)u \in \mathfrak{D}(G), \\ N_0 u &= \tfrac{1}{2}(1 - G^{-1}L)u \in \mathfrak{D}(G). \end{aligned} \tag{6.4}$$

Hence

$$P_0 u \in \mathfrak{P}_+, \qquad N_0 u \in \mathfrak{P}_-, \tag{6.5}$$

and (6.3) is proved.

Next we consider the operator F_λ, $\lambda > 1$, defined in the preceding section.

THEOREM 6.3.

$$F_\lambda \mathfrak{E}_+ \subset \mathfrak{P}_+, \qquad F_\lambda \mathfrak{E}_- \subset \mathfrak{P}_-. \tag{6.6}$$

Proof. Let $\phi \in \mathfrak{E}_+'$. Then $GF_\lambda\phi = F_\lambda G\phi = F_\lambda L\phi = LF_\lambda\phi$. Hence

$$N_0 F_\lambda \phi = \tfrac{1}{2}(1 - G^{-1}L)F_\lambda\phi = 0, \qquad \phi \in \mathfrak{E}_+'.$$

Given any $\phi \in \mathfrak{E}_+$ there exists a sequence $\phi^n \in \mathfrak{E}_+'$ with $\phi^n \to \phi$, $n \to \infty$ under the norm of $\mathfrak{E}$. But F_λ is bounded under the norm of $\mathfrak{E}$. Hence $F_\lambda\phi^n \to F_\lambda\phi$ under the norm of $\mathfrak{E}$. Since $N_0F_\lambda\phi^n = 0$, it follows that $N_0F_\lambda\phi = 0$ which proves that $F_\lambda\phi \in \mathfrak{E}_+$ for $\phi \in \mathfrak{E}_+$. Since on the other hand $F_\lambda u \in \mathfrak{P}$ for every $u \in \mathfrak{E}$ and every $1 < \lambda < \infty$ we obtain $F_\lambda\phi \in \mathfrak{P}_+$ and hence the desired inclusion. The second inclusion is proved in the analogous manner.

We will use now the operators P_0 and N_0 to define dissipative boundary conditions. This simply can be done as follows: Let I be any contraction operator mapping $\mathfrak{E}_-$ into $\mathfrak{E}_+$:

$$\begin{gathered} Iu \in \mathfrak{E}_+, \qquad u \in \mathfrak{E}_-, \\ [[Iu]] \leqq [[u]], \qquad u \in \mathfrak{E}_-. \end{gathered} \tag{6.7}$$

Then we can impose the condition

$$P_0 u = IN_0 u, \qquad u \in \mathfrak{E}. \tag{6.8}$$

Let $M(I)$ be the space of all $u \in \mathfrak{P}$ satisfying (6.8). Using this space we define an operator $M(I) = M$ in $\mathfrak{D}(M)$ by the following prescription:

$$\begin{gathered} \mathfrak{D}(M) = \{u \in \mathfrak{D}(L_1) \mid u = \phi + u_0, \quad u_0 \in \mathfrak{D}(L_0^{**}), \quad \phi \in \mathfrak{M}(I)\}, \\ Mu = L_1^{**}u, \qquad u \in \mathfrak{D}(M). \end{gathered} \tag{6.9}$$

We intend to show that, under a further condition, M in $\mathfrak{D}(M)$ is essentially maximal dissipative, i. e., its closure is maximal dissipative. In order to prove this, we first mention that the operator M is dissipative:

$$\begin{aligned} \langle M_I u, u\rangle + \langle u, M_I u\rangle &= Q(u,u) \\ &= [u, Qu] - [P_0u, P_0u] - [N_0u, N_0u] \\ &= [[IN_0u]]^2 - [[N_0u]]^2 \leqq 0 \end{aligned}$$

because of (6.7) and (6.8). Further, in order to investigate the maximality we use the Corollary of Theorem 3.1. Accordingly we have to try to prove that $(1 - L)\mathfrak{M}(I)$ is dense in $(1 - L)\mathfrak{B}$ under the norm of $\mathfrak{B}$. Now $\mathfrak{M}(I)$

consists of all elements $u \in \mathfrak{P}$ which are of the form $u = N_0 v + I N_0 v$, $v \in \mathfrak{E}$. We define the operator W by

$$W = (1 - G)(1 + G)^{-1}. \tag{6.10}$$

Obviously W is a bounded selfadjoint operator under each of the three norms $\| u \|$, $[[u]]$ and $([u])$; also W commutes with G. The norm of W is $\leqq 1$ in all three spaces $\mathfrak{B}$, $\mathfrak{E}$ and $\mathfrak{P}$.

THEOREM 6.4.

$$W\mathfrak{E}_+ \subset \mathfrak{E}_-, \qquad W\mathfrak{E}_- \subset \mathfrak{E}_+. \tag{6.11}$$

Proof. To show this first inclusion let $\phi = W\psi$, $\psi \in \mathfrak{E}_+'$; then $(1 + G)\phi = (1 - G)\psi$. But by (5.21) $G\psi = L\psi$, $\psi \in \mathfrak{E}_+'$. Hence $(1 + G)\phi = (1 - L)\psi$. Since $\frac{1}{2}(1 - L)$ is a projection, we conclude that

$$\tfrac{1}{2}(1 - L)(1 + G)\phi = (1 - L)\psi = (1 + G)\phi.$$

But by Theorem 5.2 and Theorem 6.1

$$\begin{aligned}(1 - L)N_0\phi = \tfrac{1}{2}(1 - L)(1 - G^{-1})\phi &= \tfrac{1}{2}(1 - L - LG + G)\phi \\ &= \tfrac{1}{2}(1 - L)(1 + G)\phi.\end{aligned}$$

Hence $(1 - L)N_0\phi = (1 + G)\phi$. Again we use relation (5.40) and then get $(1 + G)N_0\phi = (1 + G)\phi$. Dividing by $1 + G$ results in

$$N_0\phi = \phi, \qquad \phi \in \mathfrak{E}_-. \tag{6.12}$$

By closing the space $\mathfrak{E}_+'$ we get the same for every $\psi \in \mathfrak{E}_+$. In the analogous manner we prove the second inclusion.

Using Theorem 6.2 we see that the operator WI is a bounded transformation of $\mathfrak{E}_-$ into itself; especially:

$$[[WIu]] \leqq [[u]], \qquad u \in \mathfrak{E}_-. \tag{6.13}$$

Actually it follows that

$$[[Wu]] < [[u]] \text{ for } u \neq 0,\ u \in \mathfrak{E}, \tag{6.14}$$

and therefore in (6.13) the relation "$<$" holds unless $u = 0$. Therefore certainly the inverse $(1 + WI)^{-1}$ exists as an operator transforming a certain dense subspace of $\mathfrak{E}_-$ into $\mathfrak{E}_-$. Let us assume that $(1 + WI)^{-1}$ is bounded under the norm of $\mathfrak{E}$. This is certainly true if, for instance, the inequality (6.7) can be strengthened to

$$[[Iu]] \leqq (1 - \epsilon)[[u]], \quad u \in \mathfrak{E}_-, \quad \epsilon > 0. \tag{6.14}$$

Denote the space $(1 + WI)^{-1}\mathfrak{P}_-$ by $\mathfrak{R}(I)$. Clearly $\mathfrak{R}(I) \subset \mathfrak{E}_-$.

THEOREM 6.5. *If* $v \in \mathfrak{N}(I)$, *then*

$$u = N_0 v + I N_0 v = v + Iv \in \mathfrak{P}. \tag{6.15}$$

Proof. Let $v \in \mathfrak{N}(I)$. Then by definition of $\mathfrak{N}(I)$

$$(1 + WI)v \in \mathfrak{P}_- \subset \mathfrak{D}(G). \tag{6.16}$$

Hence $(1 - L)(1 + WI)v$ exists and by Theorem 6.1

$$(1 - L)(L + WI)v = (1 + G)(1 + WI)v. \tag{6.17}$$

Now set $u = v + Iv$ and consider $F_\lambda u$, $\lambda > 1$, with F_λ being the operator defined in (5.25). By definition of W and F_λ both operators commute. By (6.17) and Theorem 5.3 we obtain

$$F_\lambda (1 - L)(1 + WI)v = (1 - L)F_\lambda v + (1 - L)WF_\lambda Iv. \tag{6.18}$$

By Theorem 6.3 $F_\lambda v \in \mathfrak{P}_-$, $F_\lambda I \in \mathfrak{P}_+$. Hence by Theorem 6.4 $WF_\lambda Iv \in \mathfrak{P}_-$ and therefore by Theorem 6.1 and Theorem 6.2

$$(1 - L)WF_\lambda Iv = (1 + G)WF_\lambda Iv = (1 - G)F_\lambda Iv = (1 - L)F_\lambda Iv.$$

Hence (6.18) yields

$$F_\lambda (1 - L)(1 + WI)v = (1 - L)F_\lambda (v + Iv) = (1 - L)F_\lambda u. \tag{6.19}$$

Finally we observe that the following inequality holds:

$$\| w \|^2 \leqq \tfrac{1}{2} \| (1 - L)w \|^2 + [[w]]^2, \quad w \in \mathfrak{P}. \tag{6.20}$$

To prove this we simply remark that for $w \in \mathfrak{D}(G)$ the right hand side becomes

$$\begin{aligned}
\tfrac{1}{2} \| w \|^2 + \tfrac{1}{2} \| Lw \|^2 - \tfrac{1}{2}\{ (w, Lw) + (Lw, w)\} + (w, Gw) \\
= \| w \|^2 + (w, (G - L)w) \\
= \| w \|^2 + 2[w, N_0 w] \\
\geqq \| w \|^2,
\end{aligned}$$

since N_0 is positive under $[w, w]$. For $w \in \mathfrak{P}$ arbitrary we now simply choose a sequence w^n tending to w under $([u])$ and then pass to the limit. Applying (6.20) for $w = F_\lambda u$ we obtain

$$\begin{aligned}
\int_{\lambda^{-1}-0}^{\lambda} d \| E_\lambda u \|^2 = \| F_\lambda u \|^2 \leqq \tfrac{1}{2} \| (1 - L)F_\lambda u \|^2 + [[F_\lambda u]]^2 \\
= \tfrac{1}{2} \| F_\lambda (1 - L)(1 + WI)v \|^2 + [[F_\lambda u]]^2.
\end{aligned} \tag{6.21}$$

Now passing to the limit $\lambda \to \infty$ the right hand side tends to

$$\tfrac{1}{2}\,\|(1-L)(1+WI)v\|^2+[[u]]^2$$

which is a well defined number. Hence the integral $\int_0^\infty d\,\|E_\lambda u\|^2$ exists and therefore $u \in \mathfrak{B}$. Since $u \in \mathfrak{E}$ trivially holds we get $u \in \mathfrak{P}$, which proves Theorem 6.5. Now we will be able to prove the maximality of M in $\mathfrak{D}(M)$.

THEOREM 6.6. *If $(1+WI)^{-1}$ is bounded under the norm of $\mathfrak{E}$ then the operator M in $\mathfrak{D}(M)$ is essentially maximal dissipative, i.e., it closure is maximal dissipative.*

Proof. We simply observe that by Theorem 6.5 every $u=v+Iv$, $v \in \mathfrak{N}(I)$, is contained in $\mathfrak{M}(I)$. This is true because by Theorem 6.5 $u \in \mathfrak{P}$ and because $v \in \mathfrak{N}(I) \subset \mathfrak{E}$, $Iv \in \mathfrak{E}_+$ and therefore

$$P_0u=Iv=IN_0u.$$

Now for any such u we get

$$\begin{aligned}
F_\lambda(1-L)u &= (1-L)F_\lambda u=(1-L)F_\lambda v+(1-L)F_\lambda Iv\\
&=(1+G)F_\lambda v+(1-G)F_\lambda Iv \qquad (6.22)\\
&=(1+G)F_\lambda(1+WI)v=(1-L)F_\lambda(1+WI)v.
\end{aligned}$$

But by definition of $\mathfrak{N}(I)$ we get $(1+WI)\mathfrak{N}(I)=\mathfrak{P}_-$. Hence $(1+WI)v \in \mathfrak{P}_-$ and $F_\lambda(1-L)u=F_\lambda(1-L)(1+wI)v$. Passing to the limit $\lambda\to\infty$ we obtain $(1-L)u=(1-L)(1+WI)v$. Hence the space $(1-L)\mathfrak{M}(I)$ contains the space

$$(1-L)(1+WI)\mathfrak{N}(I)=(1-L)\mathfrak{P}_-$$

Consequently we only have to show that $(1-L)\mathfrak{P}_-$ is dense in $(1-L)\mathfrak{B}$ under the norm of $\mathfrak{B}$. Now let $Lf=-f$ and

$$(f,(1-L)u)=0, \qquad u \in \mathfrak{P}_-; \tag{6.23}$$

then also

$$(f,(1+L)u)=((1+L)f,u)=0, \qquad u \in \mathfrak{P}_-. \tag{6.24}$$

Hence

$$(f,Lu)=0, \qquad u \in \mathfrak{P}_-. \tag{6.25}$$

Now replace u in (6.25) by $F_\lambda u$. This is legitimate because of Theorem 6.3. Hence by Theorem 5.3

$$0=(f,LF_\lambda u)=(f,F_\lambda Lu)=-(F_\lambda f,Gu)=-[F_\lambda f,u], \qquad u \in \mathfrak{P}_-. \tag{6.26}$$

Consequently $F_\lambda f \in \mathfrak{P}_+$, $1<\lambda<\infty$. If λ tends to ∞ then $F_\lambda f$ tends to f under

the norm of $\mathfrak{B}$. But for $v \in \mathfrak{P}_+$ we obtain the estimate

$$[v, v] = (v, Gv) = (v, Lv) \leqq (v, v)$$

or $[[v]] \leqq \| v \|$, $v \in \mathfrak{P}_+$. Hence Ff converges also in $\mathfrak{P}$ and therefore $f \in \mathfrak{P}_+$. But $Lf = -f$ and therefore

$$0 \leqq [f, f] = (f, Gf) = (f, Lf) = -(f, f) \leqq 0.$$

Hence $f = 0$ and the theorem is proved.

Finally it should be remarked that obviously the same calculations can be carried through for the operator $-L_1$ instead of the operator L_1 defined in (2.1). Then we obtain the same boundary spaces, etc., but $\mathfrak{H}^0{}_+'$ and $\mathfrak{H}^0{}_-'$, $\mathfrak{E}_+$ and $\mathfrak{E}_-$, etc., get reversed. For the operators M we then get that $-M$ is an essentially maximal positive continuation of L_0. Hence the preceding theory also can be applied for getting maximal positive operators M. This and Phillips' criteria mentioned in Section 3 will imply the following

THEOREM. *If I is a unitary transformation under the norm of $\mathfrak{E}$ mapping $\mathfrak{E}_-$ onto $\mathfrak{E}_+$ and if $(1 + WI)^{-1}$ is bounded under the norm of $\mathfrak{E}$, then the operator iM in $\mathfrak{D}(M)$ is essentially selfadjoint.*

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

REFERENCES.

[1] F. E. Browder, "The eigenfunction expansion theorem for general selfadjoint singular elliptic partial differential operators, I, II," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 40 (1954), pp. 454-459, pp. 459-463.

[2] J. W. Calkin, "Abstract symmetric boundary conditions," *Transactions of the American Mathematical Society*, vol. 45 (1939), pp. 369-442.

[3] T. Carleman, "Sur la theorie mathematique de l'equation de Schrödinger," *Arkiv Matematik Astronomi Fysik*, 24B, No. 11 (1934), pp. 1-7.

[4] E. A. Coddington, "On the spectral representation of ordinary selfadjoint differential operators," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38 (1952), pp. 732-737.

[5] H. O. Cordes, "On continuation of boundary values for partial differential operators," *Pacific Journal of Mathematics*, vol. 9 (1959), pp. 987-1011.

[6] K. O. Friedrichs, "The identity of weak and strong solutions of differential operators," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 132-151.

[7] ———, "Symmetric positive linear differential equations," *Communications on Pure and Applied Mathematics*, vol. 11 1959), pp. 333-418.

[8] L. Garding, "Eigenfunction expansions connected with elliptic differential operators," *Tolfte Skandinaviska Matematikerkongresse, Lund*, 1953, pp. 44-55, 1954.

[9] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Leipzig and Berlin, 1912.

[10] L. Hörmander, "On the theory of general partial differential operators," *Acta Mathematica*, vol. 54 (1955), pp. 161-248.

[11] K. Kodaira, "On ordinary differential equations of any even order and the corresponding eigenfunction expansions," *American Journal of Mathematics*, vol. 72 (1950), pp. 502-544.

[12] N. Levinson, "The expansion theorem for singular selfadjoint linear differential operators," *Annals of Mathematics* (2), vol. 59 (1954), pp. 300-315.

[13] B. M. Levitan, "Proof of a theorem on eigenfunction expansions for selfadjoint differential equations," *Doklady Akademii Nauk USSR*, vol. 37 (1950), pp. 651-654.

[14] R. S. Phillips, "Dissipative operators and hyperbolic systems of partial differential operators," *Meddelanden fran Lunds Universitets Matematiska Seminarium* 13 (1955), pp. 1-84.

[15] ——— and P. D. Lax, to appear in *Communications on Pure and Applied Mathematics*.

[16] F. Rellich, "Störungstheorie der Spektralzerlegung II," *Mathematische Annalen*, vol. 113 (1937), pp. 677-685.

[17] L. Sandgren, "A vibration problem," *Transactions of the American Mathematical Society*, vol. 90 (1959), pp. 193-254.

[18] M. H. Stone, *Linear operators in Hilbert space*, American Mathematical Society Colloquium Publications, vol. XV, 1932.

[19] M. I. Visik, "On the general form of linear boundary problems for an elliptic differential equation," *Doklady Akademii Nauk USSR* (*N. S.*), vol. 77 (1951), pp. 373-375.

[20] ———, "On some boundary problems for elliptic differential equations," *ibid.*, vol. 77 (1951), pp. 553-555.

ON THE STIEFEL-WHITNEY CLASSES OF A MANIFOLD.* [1]

By W. S. Massey.

1. Introduction. It has been well known for many years that various relations must hold between the Stiefel-Whitney classes of the tangent bundle of a manifold which do not hold for the Stiefel-Whitney classes of an arbitrary sphere bundle. For example, Whitney [6] showed that the 3-dimensional Stiefel-Whitney class of an orientable 4-manifold is always zero. The three main theorems of this paper are results of this kind. They assert that for certain integers n and k, the k-dimensional Stiefel-Whitney class (or dual Stiefel-Whitney class) of a compact n-manifold (or a compact orientable n-manifold) is always zero.

2. Statement of results. Throughout this paper we will use only the ring of integers mod 2, Z_2, for coefficients of any cohomology groups or cohomology classes considered. The notation M^n will be consistently used to denote a compact, connected, n-dimensional manifold, $w_i \in H^i(M^n, Z_2)$ will denote the i-th Stiefel-Whitney class of its tangent bundle, and $\bar{w}_i \in H^i(M^n, Z_2)$ will denote the i-th dual Stiefel-Whitney class. The Stiefel-Whitney classes and dual Stiefel-Whitney classes are related by the following formula:

$$(2.1) \qquad (\sum_i w_i)(\sum_j \bar{w}_j) = 1.$$

According to the Whitney duality theorem, the $\bar{w}_i$ are the Stiefel-Whitney classes of the normal bundle for any differentiable imbedding of M^n in a Euclidean space of any dimension.[2]

The following three elementary properties of the Stiefel-Whitney classes of an n-manifold M^n are well known (see Wu [7]):

---

* Received February 20, 1959.

[1] During the preparation of this paper, the author was partially supported by a grant from the National Science Foundation. An abstract announcing the three main theorems of this paper was submitted to the American Mathematical Society in December, 1958 (see Notices Amer. Math. Soc., vol. 6, p. 143).

[2] In his thesis [5], R. Thom showed how to define the w_i and $\bar{w}_i$ when differentiability hypotheses are lacking.

(2.2) If n is odd, $w_n = 0$.

(2.3) $w_1 = 0$ if and only if M^n is orientable.

(2.4) For any n, $\bar{w}_n = 0$.

Our theorems extend these results.

THEOREM I. *Let M^n be a compact, n-manifold and let q be an integer such that $0 < q < n$. If $w_{n-q} \neq 0$, then there exist integers $h_1, \cdots, h_q$ such that $h_1 \geqq h_2 \geqq \cdots \geqq h_q \geqq 0$ and*

$$n = 2^{h_1} + 2^{h_2} + \cdots + 2^{h_q}.$$

Moreover, if M^n is orientable, the following additional restrictions must be imposed:

(a) $q \neq 1$,

(b) *If $n \equiv 2 \bmod 4$, then $h_q \neq 1$,*

(c) *An odd number of the h_i's are not equal to $h_q + 1$.*

The proof of this theorem will be given in § 4. For the present, we will list the following corollaries.[3]

COROLLARY 1. *If $\bar{w}_{n-1} \neq 0$, then n is a power of 2 and M^n is non-orientable.*

This is the case $q = 1$ of the theorem. Note that for n-dimensional real projective space one actually has $\bar{w}_{n-1} \neq 0$ if n is a power of 2. To prove this, one can use the determination of the Stiefel-Whitney classes of n-dimensional real projective space by E. Stiefel [4] and formula (1) above.

COROLLARY 2. *If $\bar{w}_{n-2} \neq 0$, then $n = 2^k(2^h + 1)$ for non-negative integers h and k. In addition, if M^n is orientable, the cases $n = 2(2^h + 1)$ for $h > 0$ and $n = 3.2^k$ are not possible.*

This is the case $q = 2$ of the theorem. $2^k(2^h + 1) = 2^{k+h} + 2^k$, so let $h_1 = k + h$ and $h_2 = k$. The two excluded cases correspond to cases (b) and (c) respectively of the main theorem.

COROLLARY 3. *If $n = 2^r - 1$, then $\bar{w}_i = 0$ for $i > n - r$.*

Proof. Since $2^r - 1 = 1 + 2 + 2^2 + \cdots + 2^{r-1}$, the minimum value of q which can occur in Theorem I is $q = r$.

[3] I have been informed by A. Shapiro that the result stated in corollary 1 has been obtained independently by A. Dold.

We leave it to the reader to derive other consequences of Theorem I. In doing this it is often useful to observe that the following two conditions are equivalent: (a) $n = 2^{h_1} + 2^{h_2} + \cdots + 2^{h_q}$ for non-negative integers $h_1, \cdots, h_q$. (b) In the dyadic expansion of the integer n, the digit 1 does not occur more than q times.

THEOREM II. *If n is even and M^n is orientable, then $w_{n-1} = 0$.*

Wu indicates a proof of this result in case $n \equiv 2 \bmod 4$ (see [7]). The proof for the case $n \equiv 0 \bmod 4$ if given in § 5.

THEOREM III. *If $n \equiv 3 \bmod 4$ and M^n is orientable, then $w_n = w_{n-1} = w_{n-2} = 0$.*

This theorem is an easy consequence of Wu's formulas [7]. The proof is given in § 3.

In a certain sense, Theorems II and III together with statements (2.2) and (2.3) are the best that one can hope for in this direction. This may be seen by consideration of certain examples, as follows:[4]

In the case of non-orientable manifolds, statement (2.2) above is the best possible. For if n is even, $n = 2k$, then $M^n = (P_2)^k$ (the Cartesian product of k copies of the real projective plane, P_2) has non-vanishing Stiefel-Whitney classes in all dimensions, while if n is odd, $n = 2k + 1$, then $M^n = (P_2)^k \times S^1$ (where S^1 denotes a 1-sphere) has non-vanishing Stiefel-Whitney classes in all dimensions $< n$.

The case of orientable manifolds is more complicated. First consider the case where $n = 4k$. Let $P_2(C)$ denote the complex projective plane (a 4-dimensional manifold), and let $M^n = [P_2(C)]^k$, the Cartesian product of k copies. Then $w_i \neq 0$ for all even integers $i \leq n$; in particular, $w_{n-2} \neq 0$, so Theorem II can not be improved if $n = 4k$. If $n = 4k + 1$, one may obtain examples of an M^n for which $w_{n-1} \neq 0$ by taking $M^n = [P_2(C)]^k \times S^1$, or $M^n = P(1, 2k)$, a manifold considered by A. Dold in [1]. For $n = 4k + 2$, one may obtain an example of an M^n for which $w_{n-2} \neq 0$ by taking M^n

[4] For the proof of the assertions made in the following paragraphs about these example, the following result is needed. Let M and M' be compact manifolds. Identify the cohomology ring of the product space, $H^*(M \times M', Z_2)$, with the tensor product $H^*(M, Z_2) \otimes H^*(M', Z_2)$ as usual. If $w = 1 + w_1 + w_2 + \cdots$ and $w' = 1 + w'_1 + w'_2 + \cdots$ denote the total Stiefel-Whitney classes of M and M' respectively, then $w \otimes w'$ is the total Stiefel-Whitney class of $M \times M'$. For the proof, see Thom, [5], pp. 142-143. One also needs to know that for the real projective plane, P_2, $w_1 \neq 0$ and $w_2 \neq 0$ (see [4]); for the circle, S^1, $w_1 = 0$; and for the complex projective plane, $P_2(C)$, $w_2 \neq 0$ and $w_4 \neq 0$. The Stiefel-Whitney classes of Dold's manifolds $P(m, n)$ have been computed by Dold [1].

$= [P_2(C)]^k \times S^1 \times S^1$ or $M^n = P(1, 2k) \times S^1$; therefore Theorem II can not be improved in this case either. Similarly, for $n = 4k + 3$ one obtains examples where $w_{n-3} \neq 0$ by taking $M^n = [P_2(C)]^k \times [S^1]^3$ or $M^n = P(1, 2k) \times S^1 \times S^1$. Thus Theorem III can not be improved. Whether or not Theorem I is a best possible theorem in this sense seems like a much more difficult question.

Of course there are other directions in which one could try to extend these theorems. For example, one could try to determine more general kinds of relations between Stiefel-Whitney classes of a manifold. An example is the relation $w_1 w_2 = 0$ which holds in all manifolds of dimension $\leqq 5$ (see Wu [7]). This important problem seems very difficult, and outside of the case considered by A. Dold in [2], very little is known about it. One of the most pertinent problems in this connection is the following: Can any relation which holds between the Stiefel-Whitney classes of every n-manifold (or every orientable n-manifold) be derived from the formulas of Wu ([7] and [8])?

It should be pointed out that Theorem I may have implications for the problem of determining the lowest dimensional Euclidean space in which it is possible to imbed a given manifold. Whitney has proved that it is possible to imbed any n-dimensional smooth manifold differentiably in $2n$-dimensional Euclidean space. Moreover, if $n = 2^k$, then it is possible to give an example of an n-manifold which can not be imbedded in Euclidean $(2n-1)$-space: n-dimnsional real projective space P_n is such an example. To prove that P_n can not be imbedded in Euclidean $(2n-1)$-space if $n = 2^k$ one uses the fact that $\bar{w}_{n-1} \neq 0$. On the other hand, Corollary 1 of Theorem I shows that $\bar{w}_{n-1} = 0$ for any n-manifold if n is not a power of 2. Thus it is natural to ask the following question: If n is not a power of 2, can any n-manifold be imbedded in Euclidean $(2n-1)$-space? If the answer to this question is "no," it will require new methods to prove the existence of a counter-example.[5]

3. Notation and preliminary results. We will use the following notation and ideas in what follows. They are due to W. T. Wu [7].

(a) $U_i \in H^i(M^n, Z_2)$ is the unique cohomology class such that

$$x \cdot U_i = Sq^i(x) \tag{3.1}$$

for any $x \in H^{n-i}(M^n, Z_2)$. The existence and uniqueness of the U_i follow from the Poincaré duality theorem. Note that $U_0 = 1$, and $U_i = 0$ if $i > \frac{1}{2}n$.

[5] As a matter of fact, it was the search for examples of n-manifolds with $\bar{w}_{n-1} \neq 0$ which led the author to the discovery of Theorem I.

(b) The cohomology classes $\bar{U}_i \in H^i(M^n, Z_2)$ are defined inductively by the equation

$$(\sum_i U_i) \cdot (\sum_i \bar{U}_i) = 1. \tag{3.2}$$

Here again $\bar{U}_0 = 1$. However it is *not* true in general that $\bar{U}_i = 0$ for $i > \frac{1}{2}n$.

Wu proved the Stiefel-Whitney classes and dual Stiefel-Whitney classes may be expressed in terms of the U_i and $\bar{U}_i$ respectively as follows:

$$w_k = \sum_i Sq^{k-i} U_i, \tag{3.3}$$

$$\bar{w}_k = \sum_i Sq^{k-i} \bar{U}_i. \tag{3.4}$$

These formulas are basic for all later computations.

In the following lemmas we record for later use some well known facts.

LEMMA 1. *A compact n-manifold, M^n, is orientable if and only if the homomorphism $Sq^1 : H^{n-1}(M^n, Z_2) \to H^n(M^n, Z_2)$ is trivial.*

This lemma is easily proved by using the fact that the homomorphism Sq^1 is composition of the Bockstein homomorphism together with reduction mod 2, plus the known structure of the integral cohomology group in dimension n of an n-manifold.

LEMMA 2. *If M^n is orientable, then $Sq^i : H^{n-i}(M^n, Z_2) \to H^n(M^n, Z_2)$ is zero for i odd.*

This follows from the known fact that $Sq^i = Sq^1 Sq^{i-1}$ for i odd, together with Lemma 1.

LEMMA 3. *If M^n is orientable, then $U_i = \bar{U}_i = 0$ for i odd.*

The fact that $U_i = 0$ for i odd follows from Lemma 2 and the definition of U_i. Then one can prove that $\bar{U}_i = 0$ for i odd by using formula (3.2).

In our proofs we need to make use of known properties of Steenrod squares and iterated Steenrod squares. For the sake of convenience, we will use the terminology and notation of Serre [3]. We assume the reader is familar with the properties of Steenrod squares as listed in § 2 of Serre's paper. Especially frequent use will be made of the properties of the homomorphism Sq^1. According to Cartan's formula,

$$Sq^1(x \cdot y) = (Sq^1 x) \cdot y + x \cdot (Sq^1 y), \tag{3.5}$$

i.e., Sq^1 is a derivation of the algebra $H^*(X, Z_2)$. In particular,

$$Sq^1(x^k) = k x^{k-1} \cdot Sq^1 x \tag{3.6}$$

for any positive integer k. Note also that

(3.7) $$Sq^1Sq^1 = 0.$$

This implies that for any odd integer i,

(3.8) $$Sq^1Sq^i = 0.$$

We conclude this section by proving Theorem II. For an orientable manifold M^n of dimension $n = 4k + 3$, $U_i = 0$ unless i is even and $0 \leqq i \leqq 2k$ (see Lemma 3). From this and (3.3) it follows that $w_i = 0$ for $i > 4k$ as desired.

4. Proof of the Theorem I. In the proof of Theorem I, frequent use will be made of the properties of iterated Steenrod squares. If $I = (i_1, i_2, \cdots, i_r)$ is any sequence of positive integers, then the notation Sq^I denotes the iterated Steenrod square $Sq^{i_1}Sq^{i_2} \cdots Sq^{i_r}$. Such a sequence $I = (i_1, i_2, \cdots, i_r)$ is *admissible* if $i_1 \geqq 2i_2, i_2 \geqq 2i_3, \cdots, i_{r-1} \geqq 2i_r$. Every iterated Steenrod square may be expressed as a sum of admissible iterated Steenrod squares by repeated use of Adem's relations (see Serre [3], § 32).

With any admissible sequence of positive integers $I = (i_1, i_2, \cdots, i_r)$ one may associate a sequence of non-negative integers $(\alpha_1, \alpha_2, \cdots, \alpha_r)$ by the formulas

(4.1) $$\alpha_1 = i_1 - 2i_2, \alpha_2 = i_2 - 2i_3, \cdots, \alpha_{r-1} = i_{r-1} - 2i_r, \alpha_r = i_r.$$

It is clear that the sequence $(\alpha_1, \cdots, \alpha_r)$ determines without ambiguity the sequence $(i_1, \cdots, i_r)$. The integer $n(I) = i_1 + \cdots + i_r$ is called the *degree* of I, and $e(I) = \alpha_1 + \cdots + \alpha_r$ is called the *excess* of I.

LEMMA 4. *For any* mod 2 *cohomology class x, $Sq^I(x) = 0$ if the degree of x is less than the excess of I.*

The proof depends on the fact that $Sq^k(x) = 0$ if k is greater than the degree of x. The details are left to the reader.

LEMMA 5. *Let $I = (i_1, \cdots, i_r)$ be an admissible sequence of excess $e(I)$. Then there exists a unique admissible sequence $J = (j_1, \cdots, j_s)$ and a power of 2, $m = 2^k$, such that for any cohomology class x of degree $e(I)$,*

$$Sq^I(x) = (Sq^Jx)^m$$

and $e(J) < e(I)$.

For the proof, see the proof of Lemma 1, p. 204, of Serre [3].

Lemmas 4 and 5 together show that when considering iterated Steenrod squares operating on cohomology classes x of a fixed degree q, we can restrict our attention to those iterated squares Sq^I such that $e(I) \leqq q-1$. In this case it is convenient (following Serre [3], p. 212) to let $\alpha_0 = q-1-e(I)$. Then one can derive the following formulae in case x is any mod 2 cohomology class of degree q:

$$
\begin{aligned}
&\text{degree}\,(Sq^I x) = n(I) + q \\
(4.2)\qquad &= \sum_{i=1}^{r} (2^i - 1)\alpha_i + q = \sum_{i=1}^{r} 2^i\alpha_i - e(I) + q \\
&= \sum_{i=1}^{r} 2^i\alpha_i + \alpha_0 + 1 = 1 + \sum_{i=0}^{r} 2^i\alpha_i.
\end{aligned}
$$

Since $\sum_{i=0}^{r}\alpha_i = \alpha_0 + e(I) = q-1$, there are in all $(q-1)$ powers of 2 in formula (4.2). Therefore we can rewrite this formula as follows,

$$
(4.3)\qquad \text{degree}\,(Sq^I x) = 1 + 2^{h_1} + 2^{h_2} + \cdots + 2^{h_{q-1}},
$$

where $h_1 \geqq h_2 \geqq \cdots \geqq h_{q-1} \geqq 0$, and 2^i occurs α_i times in this sum (this is formula (17.5) of Serre [3], p. 212).

Next we will prove a couple of lemmas which are needed in the proof of Theorem I.

Lemma 6. *For any* $x \in H^k(M^n, Z_2)$, $0 < k < n$, $x\cdot\bar{w}_{n-k} = \sum_{i>0}(Sq^i x)\bar{w}_{n-k-i}$.

Proof. By equation (3.4),

$$
\begin{aligned}
\bar{w}_{n-k} &= \sum_{i\geqq 0} Sq^i\bar{U}_{n-k-i} \\
&= \bar{U}_{n-k} + \sum_{i>0} Sq^i\bar{U}_{n-k-i}.
\end{aligned}
$$

By equation (3.2),

$$
\bar{U}_{n-k} = \sum_{i>0} U_i\bar{U}_{n-k-i},
$$

hence

$$
\bar{w}_{n-k} = \sum_{i>0} (U_i\bar{U}_{n-k-i} + Sq^i\bar{U}_{n-k-i}).
$$

Therefore if $x \in H^k(M^n, Z_2)$,

$$
x\cdot\bar{w}_{n-k} = \sum_{i>0} (x\cdot U_i\bar{U}_{n-k-i} + x\cdot Sq^i\bar{U}_{n-k-i}).
$$

But

$$
\begin{aligned}
xU_i\bar{U}_{n-k-i} &= U_i(x\bar{U}_{n-k-i}) = Sq^i{}_{n-k-i})\,(x\bar{U}_{n-k-i}) \\
&= \sum_{r=0}^{i} (Sq^r x)\,(Sq^{i-r}\bar{U}_{n-k-i}),
\end{aligned}
$$

from which it follows that

$$\begin{aligned} x\cdot\bar{w}_{n-k} &= \sum_{i>0}\sum_{r=1}^{i}(Sq^r x)(Sq^{i-r}\bar{U}_{n-k-i}) \\ &= \sum_{0<r\leqq i}(Sq^r x)(Sq^{i-r}\bar{U}_{n-k-i}) \\ &= \sum_{r>0}[(Sq^r x)\sum_{j\geqq 0}Sq^j\bar{U}_{n-k-r-j}] \\ &= \sum_{r>0}(Sq^r x)\bar{w}_{n-k-r}. \end{aligned}$$

as was to be proved.

LEMMA 7. *The homomorphism $H^k(M^n, Z_2)\to H^n(M^n, Z_2)$ defined by $x\to x\cdot\bar{w}_{n-k}$ is a sum of iterated Steenrod squares.*

In view of Lemma 6, this lemma is obvious: one applies Lemma 6 repeatedly until the desired reduction to a sum of iterated Steenrod squares is obtained.

We are now in a position to prove Theorem I. Assume that $\bar{w}_{n-q}\in H^{n-q}(M^n, Z_2)$ is non-zero. By the Poincaré duality theorem, the homomorphism $H^q(M^n, Z_2)\to H^n(M^n, Z_2)$ defined by $x\to x\cdot\bar{w}_{n-q}$ is also non-zero. By Lemma 7, this homomorphism is a sum of iterated Steenrod squares, which we may assume to be admissible on account of Adem's relations. Hence the hypothesis of the theorem implies the following statement: There exists a non-zero admissible iterated Steenrod square

$$Sq^I\colon H^q(M^n, Z_2)\to H^n(M^n, Z_2),$$

where $I=(i_1,\cdots,i_r)$. By Lemma 4, $e(I)\leqq q$. Moreover, if $e(I)=q$, it follows from Lemma 5 that there exists an admissible sequence $J=(j_1,\cdots,j_s)$ and a power of 2, $m=2^k$, such that

$$Sq^I(x)=[Sq^J(x)]^m$$

and $e(J)<q$. Therefore

$$(4.4)\qquad \begin{aligned} n &= \operatorname{degree}(Sq^I x)=2^k\cdot\operatorname{degree}(Sq^J x) \\ &= 2^k(2^{k_1}+2^{k_2}+\cdots+2^{k_{q-1}}+1) \end{aligned}$$

by equation (4.3). Here $k_1, k_2,\cdots$ are integers such that $k_1\geqq k_2\geqq\cdots\geqq k_{q-1}\geqq 0$. If now we let

$$(4.5)\qquad h_1=k_1+k, h_2=k_2+k,\cdots,h_{q-1}=k_{q-1}+k, h_q=k,$$

then (4.4) takes the form

$$n = 2^{h_1} + 2^{h_2} + \cdots + 2^{h_q} \tag{4.6}$$

with $h_1 \geqq h_2 \geqq \cdots \geqq h_q \geqq 0$, and the first part of the theorem is proved.

Next, we will assume that M^n is orientable and prove the remaining parts of the theorem. In this case we can apply the results of Lemmas 1 and 2.

First assume that $q = 1$. Then $n = 2^{h_1}$ from what we have just proved, and $h_1 = k$ according to equation (4.5). Therefore the only non-zero iterated Steenrod square

$$Sq^I : H^1(M^n, Z_2) \to H^n(M^n, Z_2)$$

would be of the form $Sq^I(x) = x^n$ with $n = 2^k$. Since n is even, $x^n = Sq^1(x^{n-1})$ by equation (3.6). By use of Lemma 1, we see that if $x^n \neq 0$, then M^n is non-orientable, as was to be proved.

Next we will consider the case were $n \equiv 2 \bmod 4$, i.e., $n = 4l + 2$, and $h_q = 1$. Then it follows from (4.5) that $k = 1$. Therefore $Sq^I(x) = [Sq^J(x)]^2$; and by equation (4.4),

$$n = \text{degree}(Sq^I x) = 2 \cdot \text{degree}(Sq^J x).$$

Hence degree$(Sq^J x) = n/2 = 2l + 1$. Thus

$$Sq^I(x) = Sq^{2l+1}[Sq^J(x)]$$

which is zero by Lemma 2. But this is a contradiction. Thus part (b) of Theorem I is proved.

Finally, we consider the case where an odd number of the h_i's are equal to $h_q + 1$. Then it follows from equation (4.5) that an odd number of the k_i's are equal to 1. Thus in (4.3), the summand 2^1 occurs an odd number of times,[6] i.e., α_1 is odd, it follows from equation (4.1) that j_1 is odd in the expression

$$Sq^I(x) = [Sq^J(x)]^m = [Sq^{j_1} \cdots Sq^{j_s}(x)]^m,$$

where $m = 2^k$. Since j_1 is odd,

$$Sq^{j_1} = Sq^1 Sq^{j_1 - 1},$$

and $Sq^1 Sq^J(x) = 0$ by equation (3.8). Therefore

$$[Sq^J(x)]^m = Sq^1\{[Sq^{j_1-1} Sq^{j_2} \cdots Sq^{j_s}(x)] \cdot [Sq^J(x)]^{m-1}\}$$

which is again zero by Lemma 1. Thus we have again reached a contradiction, and part (c) is proved.

---

[6] Actually, we are here concerned with the analog of equation (4.3) which is obtained by replacing I by J and h_i by k_i for $i = 1, 2, \cdots, q - 1$.

5. Proof of Theorem II for the case $n \equiv 0 \bmod 4$. The following well-known lemma will be used in the course of the proof:

LEMMA 8. *If x is a* mod 2 *cohomology class of degree* 1, *then*

$$Sq^j(x^k) = C_j^k x^{k+j},$$

where C_j^k is the binomial coefficient reduced mod 2. *In particular, if k is a power of 2, then $Sq^j x^k = 0$ unless $j = 0$ or $j = k$.*

The proof is left to the reader.

Now assume that M^n is a compact orientable manifold of dimension $n = 4k$. Then

$$w_{n-1} = w_{4k-1} = Sq^{2k-1}U_{2k}$$

by (3.3). To prove that $w_{n-1} = 0$, it suffices to prove that $x \cdot w_{n-1} = 0$ for any $x \in H^1(M^n, Z)$. Now

$$\begin{aligned} x \cdot w_{n-1} &= x \cdot Sq^{2k-1}U_{2k} \\ &= Sq^{2k-1}(x \cdot U_{2k}) + (Sq^1 x)(Sq^{2k-2}U_{2k}). \end{aligned}$$

However the first term on the right is zero by Lemma 2, and in the second term, $Sq^1 x = x^2$. Therefore

$$x \cdot w_{n-1} = x^2 \cdot Sq^{2k-2}U_{2k} \tag{5.1}$$

We will now show that if $p = 2^q$ is a power of 2 and $2 \leqq p < 2k$, then

$$x^p \cdot Sq^{2k-p}U_{2k} = x^{2p} \cdot Sq^{2k-2p}U_{2k}. \tag{5.2}$$

To prove this, one computes as follows:

$$x^p Sq^{2k-p}\, U_{2k} = Sq^{2k-p}(x^p \cdot U_{2k}) + x^{2p} Sq^{2k-2p}\, U_{2k}. \tag{5.3}$$

Here we have used the formula for the Steenrod square of a cup product together with Lemma 8. Next, note that

$$\begin{aligned} Sq^{2k-p}(x^p\, U_{2k}) &= U_{2k-p}(x^p\, U^{2k}) \\ &= U_{2k}(x^p\, U_{2k-p}) = Sq^{2k}(x^p\, U_{2k-p}) \\ &= (x^p\, U_{2k-p})^2 = x^{2p} \cdot U^2{}_{2k-p} \\ &= Sq^1(x^{2p-1} \cdot U^2{}_{2k-p}) = 0 \end{aligned} \tag{5.4}$$

by (3.5), (3.6), and Lemma 1. Substitution of (5.4) in (5.3) gives (5.2), as desired.

One can now apply (5.2) to (5.1) repeatedly with $p = 2, 4, 8, \cdots$, in

succession. If n is not a power of 2, this procedure leads to the result that $x \cdot w_{n-1} = 0$, as desired. If n is a power of 2, this same procedure shows that $x \cdot w_{n-1} = x^n$ for any $x \in H^1(M^n, Z)$. However in this case, since n is even,

$$x^n = Sq^1(x^{n-1})$$

by (3.6). But $Sq^1(x^{n-1}) = 0$ by Lemma 1, as was to be proved. The proof of Theorem II is complete.

BROWN UNIVERSITY.

REFERENCES.

[1] A. Dold, "Erzeugende der Thomschen Algebra $\mathfrak{N}$," *Mathematische Zeitschrift*, vol. 65 (1956), pp. 25-35.

[2] ———, "Vollständigkeit der Wuschen Relationen Zwischen der Stiefel-Whitneyschen Zahlen Differenzierbaren Mannigfaltigkeiten," *ibid.*, vol. 65 (1956), pp. 200-206.

[3] J. P. Serre, "Cohomologie Modulo 2 des Complexes d'Eilenberg-MacLane," *Commentarii Mathematicii Helvetici*, vol. 27 (1953), pp. 198-232.

[4] E. Stiefel, "Über Richtungsfelder in den Projektiven Räumen und einen Satz aus der Reelen Algebra," *ibid.*, vol. 13 (1940), pp. 201-218.

[5] R. Thom, "Espaces Fibrés en Spheres et Carrés de Steenrod," *Annales Scientifiques de l'Ecole Normale Supérieure*, vol. 69 (1952), pp. 109-182.

[6] H. Whitney, "On the topology of differentiable manifolds," *Lectures in Topology*, University of Michigan Press, 1941, pp. 101-141.

[7] W. T. Wu, "Classes Caractéristiques et i-Carrés d'une Variété," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 230 (1950), pp. 508-511.

[8] ———, "On squares in Grassmannian manifolds," *Acta Scientia Sinica*, vol. 2 (1953), pp. 91-115 (these results were first announced in *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 230 (1950), pp. 918-920).

ON A SUBALGEBRA OF $L(-\infty, \infty)$.*

By JOHN WERMER.

Let L denote the group algebra of the real line, i.e. the algebra under convolution of summable functions on $(-\infty, \infty)$ with norm defined by

$$\| f \| = \int_{-\infty}^{\infty} | f(x) | \, dx.$$

Let L^+ be the closed subalgebra of L consisting of all functions in L which vanish on the negative half-line.

In his paper "On the Maximality of Vanishing Algebras," [3], A. B. Simon refers to the following result:

THEOREM. *L^+ is a maximal closed subalgebra of L.*
We now give a proof of this theorem, by deducing it from the following:

LEMMA. *Let B be a commutative semi-simple Banach algebra with maximal ideal space the unit circle $|\lambda| = 1$, hence a function-algebra on $|\lambda| = 1$. Assume:*

(1) *The functions λ and $1/\lambda$ lie in B and generate a dense subalgebra of B.*

Let B^+ be the algebra of those functions in B which have continuous extensions to $|\lambda| \leqq 1$ which are analytic in $|\lambda| < 1$. Then B^+ is a maximal subalgebra of B.

Proof. B^+ contains all powers of λ with non-negative exponent. Let Q be any closed proper subalgebra of B containing B^+. Then $1/\lambda \notin Q$, for else $\lambda^n \in Q$ for all integers n, whence $Q = B$ by (1). Hence there exists a multiplicative linear functional χ on the algebra Q with $\chi(\lambda) = 0$. For all f in Q, then

$$| \chi(f) | \leqq \lim_{n \to \infty} \| f^n \|_Q^{1/n} = \lim_{n \to \infty} \| f^n \|_B^{1/n} = \max_{|\lambda|=1} | f(\lambda) |.$$

By the Riesz representation theorem, there exists a measure $d\mu$ on $|\lambda| = 1$,

* Received February 24, 1959.

$d\mu \not\equiv 0$, with

$$\chi(f) = \int_{|\lambda|=1} f(\lambda)\, d\mu(\lambda), \text{ all } f \text{ in } Q.$$

In particular, if $g \in Q$, $\lambda^n \cdot g^m \cdot \lambda \in Q$ if $n \geqq 0$, $m \geqq 0$, and $\chi(\lambda^n g^m \lambda) = 0$, since $\chi(\lambda) = 0$, whence

$$0 = \int_{|\lambda|=1} g^m(\lambda) \lambda^n \, d\sigma(\lambda), n, m \geqq 0,$$

where $d\sigma(\lambda) = \lambda\, d\mu(\lambda) \not\equiv 0$. It follows from this, by a theorem given in [1], that g admits an analytic extension to $|\lambda| < 1$. Since this holds for all g in Q, $Q = B^+$. This proves the Lemma.

Note. This Lemma was suggested to the author by the reasoning of Singer and Hoffman, in their paper, [2]. The same Lemma was independently noticed by DeLeeuw.

Proof of Theorem. We prove the maximality of L^+ by introducing a related algebra B on the unit circle which satisfies the conditions of the Lemma. To each $f \in L$ assign f^* on $|\lambda| = 1$ defined by

$$f^*(\lambda) = \int_{-\infty}^{\infty} f(x) \exp(-x(1+\lambda)/(1-\lambda))\, dx.$$

Put $B = \{f^* + c \mid f \in L,\ c \text{ a constant}\}$ and norm B by setting $\| f^* + c \|_B = \|f\|_L + |c|$. Then B is a commutative semi-simple Banach algebra, since L is, and also $1 \in B$. Also the maximal ideal space of B is the unit circle. We claim λ and $1/\lambda \in B$. For let $f_1(x) = 0$, $x < 0$, $f_1(x) = 2e^{-x}$, $x \geqq 0$. Then $f_1 \in L$ and so $f_1^* \in B$. But $f_1^* = 1 - \lambda$ and so $\lambda \in B$. Similarly, $1/\lambda \in B$. We next claim $\{\lambda^n \mid -\infty < n < \infty\}$ span B. For let Φ be any bounded linear functional on B. By the well-known representation of such functionals on L, there exists a bounded function ϕ on $(-\infty, \infty)$ with

$$\Phi(f^*) = \int_{-\infty}^{\infty} f(x) \phi(x)\, dx, \text{ all } f \in L.$$

Let f_m be the m-fold convolution of f_1. Then $f_m(x) = 0$, $x < 0$, $f_m(x) = c_m x^{m-1} e^{-x}$, $x \geqq 0$, c_m a constant. Let Φ be a bounded linear functional on B with $\Phi(\lambda^n) = 0$, $n = 0, \pm 1, \pm 2, \cdots$. Then

$$0 = \Phi((f_1^*)^m) = \Phi(f_m^*) = \int_{-\infty}^{\infty} f_m(x) \phi(x)\, dx, \ m \geqq 0$$

or

$$0 = \int_0^{\infty} x^{m-1} e^{-x} \phi(x)\, dx, \ m = 1, 2, \cdots.$$

Hence $\phi(x)\equiv 0$, $x>0$. Similarly $\phi(x)\equiv 0$, $x<0$. Thus $\Phi\equiv 0$ and so $\{\lambda^n \mid -\infty<n<\infty\}$ spans B as claimed. Let now B^+ be the algebra of functions in B admitting analytic extensions to $|\lambda|<1$. Since B satisfies (1), as we have just seen, B^+ is maximal in B by the Lemma.

We now claim that $B^+=\{f^*+c \mid f\in L^+,\ c \text{ constant}\}$. If $f\in L$, the function F defined by

$$F(s)=\int_{-\infty}^{\infty} f(x)e^{-sx}\,dx,\quad s=it,\ t \text{ real}$$

is related to f^* by $f^*(\lambda)=F((1+\lambda)/(1-\lambda))$. Hence $f^*+c\in B^+$ if and only if F admits an analytic extension to the right half-plane, continuous in the closed right half-plane, including ∞. It is clear that this occurs if and only if $f\in L^+$. Thus $B^+=\{f^*+c \mid f\in L^+,\ c \text{ constant}\}$, as asserted. Let K be any closed proper subalgebra of L containing L^+. Put $K^*=\{f^*+c \mid f\in K,$ c constant$\}$. Then K^* is a closed proper subalgebra of B, and since $K\supseteq L^+$, $K^*\supseteq\{f^*+c \mid f\in L^+,\ c \text{ constant}\}$ which is B^+, as we just saw. Since B^+ is maximal, $K^*=B^+$, and so $K=L^+$. This proves the theorem.

BROWN UNIVERSITY.

REFERENCES.

[1] J. Wermer, "On algebras of continuous functions," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 866-869.

[2] K. Hoffman and J. M. Singer, "Maximal subalgebras of $C(\Gamma)$," *American Journal of Mathematics*, vol. 79, No. 2 (1957), pp. 295-305.

[3] A. B. Simon, "On the maximality of vanishing algebras," *American Journal of Mathematics*, vol, 81, No. 3 (1959), pp. 613-616.

LIMIT PROPERTIES AT ZERO OF THE MARKOV SEMI-GROUP.*

By RAFAEL V. CHACON.[1]

Introduction. Let $(\Omega, \mathfrak{F}, P)$ be a probability triple, and $\mathcal{B}$ be a Borel field of sets of points of X, and $\{x(t,\omega), t \in (0, +\infty)\}$ be a Markov process with state space X, and with (stationary) transition (probability) function $P_t(x, A)$. In other words, there exists a function $P_t(x, A)$, defined for $t \in (0, +\infty)$, $x \in X$, and $A \in \mathcal{B}$ such that

$$P\{x(t+s) \in A \mid x(s,\omega)\} = P_t(x(s,\omega), A), \text{p. 1.}$$

We suppose further that the transition function satisfies:

(0.1) For each $t \in (0, +\infty), x \in X, P_t(x, \cdot)$ is a probability measure on $\mathcal{B}$.

(0.2) For each $t \in (0, +\infty), A \in \mathcal{B}, P_t(\cdot, A)$ is $\mathcal{B}$-measurable.

(0.3) The Chapman-Kolmogorov equation

$$P_{t+s}(x, A) = \int_X P_t(y, A) P_s(x, dy).$$

An additional condition is often imposed as well:

$$\lim_{t \to 0} P_t(x, A) = 1 \quad \text{if } x \in A. \tag{0.4'}$$

We do not impose hypothesis (0.4'); our goal is to investigate to what extent this condition is satisfied under some minimal additional assumptions. In an earlier paper [1] it has been shown that if a condition like (0.4') is satisfied, then the transition functions will be continuous in t. In view of this fact, we suppose that the transition functions satisfy also:

(0.4) For each $x \in X$ and $A \in \mathcal{B}$, $P_t(x, A)$ is a continuous function of t.

If we suppose that X is the set of positive integers, and that $\mathcal{B}$ is the Borel field of subsets of X, then the process is called a Markov chain. Doob [3] has proved the following theorem:

* Received February 24, 1959.

[1] This work was supported by the ONR contract at Cornell University, No. Nonr-401(03) and by the U. S. Army Research Center at the University of Wisconsin.

THEOREM 0.1. *A Markov chain has continuous transition functions if and only if the integers may be divided into pairwise disjoint sets $F, I_1, I_2, \cdots$, such that*

$$\text{(i)} \qquad P_t(i, \{j\}) = 0, \; j \in F,$$

$$\text{(ii)} \qquad \lim_{t\to 0} P_t(i, I_j) = \delta_{I_j}(i), \; i \in \bigcup_{j=1}^{\infty} I_j,$$

$$\text{(iii)} \qquad P_t(i_1, \{j\}) = u_j P_t(i_2, I_k), \; j \in I_k, \; i_1, i_2 \in I_l,$$

$$\text{(iv)} \qquad P_t(i, \{j\}) = u_j P_t(i, I_k), \; j \in I_k, \; i \in F,$$

where $\{u_j\}$ is a sequence of non-negative numbers, and

$$\text{(v)} \qquad \lim_{t\to 0} P_t(i, \{j\}) \text{ exists}, \; i, j = 1, 2, \cdots.$$

We define the function δ as follows:

$$\delta_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

We establish a result analogous to Theorem 0.1, assuming that the transition probabilities satisfy conditions (0.1), (0.2), (0.3) and (0.4). In the special case considered in Theorem 0.1, the function U which we obtain can be further analyzed, and the details of Theorem 0.1 can thereby be obtained. We remark that the method of proof is different from Doob's and that his proof doesn't seem to generalize.

It is possible to define certain semi-groups, using the transition functions, on suitably chosen Banach spaces. However, Hille's work on the existence of the identity as a limit as the parameter tends to zero does not yield our result in any obvious way.

1. Results and proofs. We suppose in what follows that we have a Markov process whose transition function satisfies (0.1), (0.2), (0.3), and (0.4). Blackwell [2] gave a special case of Lemma 1.1 to study idempotent Markov chains.

LEMMA 1.1. *If A and B are two sets in $\mathcal{B}$, then (where $\psi_C(\cdot)$ stands for the characteristic function of the set C)*

$$\int_B \{ \int_{cB} P_s(z, A) P_t(y, dz) \} P_u(x, dy)$$
$$= \int \psi_{cB}(z) P_s(z, A) P_{t+u}(x, dz) - \int_{cB} \{ \int_{cB} P_s(z, A) P_t(y, dz) \} P_u(x, dy),$$

and

$$\int_{cB}\{\int_B P_s(z,A)P_t(y,dz)\}P_u(x,dy)$$
$$=\int \psi_{cB}(y)P_{s+t}(y,A)P_u(x,dy)-\int_{cB}\{\int_{cB}P_s(z,A)P_t(y,dz)\}P_u(x,dy).$$

Proof. Follows at once by transposing the second terms on the right to the left and combining each of the terms into single integrals. That the interchange of order of integration is valid can be easily shown.

LEMMA 1.2. *If A and B are two sets in $\mathfrak{B}$, and if a is a real number, and $s>0$, $t>0$, $u>0$,*

$$\lim_{t\to 0}\{\int_B[\int_{cB}(P_s(z,A)P_t(y,dz)]P_u(x,dy)$$
$$-\int_{cB}[\int_B(P_s(z,A)-a)P_t(y,dz)]P_u(x,dy)\}=0.$$

Proof. Follows at once from Lemma 1.1.

LEMMA 1.3. *For each $s>0$, $u>0$, c real, $x\in X$ and $A\in\mathfrak{B}$, fixed, if $C=\{z\colon P_s(z,A)<c\}$ then*

$$\lim_{t\to 0}P_t(y,C)=\delta_C(y)\ \textit{in}\ P_u(x,\cdot)\textit{-measure.}$$

The lemma remains valid if "$<$" is replaced by either "$\leqq$" or "$>$" or "$\geqq$" in the definition of the set C.

Proof. Using Lemma 1.2, put $B=B_n=\{z\colon P_s(z,A)>c-1/n\}$ and $a=a_n=c-1/n$. Both terms have the same sign in this application, and hence each must have a zero limit. We have, using the second term,

$$\lim_{t\downarrow 0}\int\psi_{cB_n}(y)[\int\psi_{B_n}(z)\{P_s(z,A)-c+1/n\}P_t(y,dz)]P_u(x,dy)=0.$$

Since the inner integral is of constant sign as a function of y, we have that

$$\psi_{cB_n}(y)\int\psi_{B_n}(z)\{P_s(z,A)-c+1/n\}P_t(y,dz)$$

converges in $P_u(x,\cdot)$-measure to zero as $t\to 0$. From this it follows that

$$\psi_{cB_n}(y)P_t(y,\{z\colon P_s(z,A)\geqq c\})$$

also converges in $P_u(x,\cdot)$-measure to zero as $t\to 0$, and from this that

$$\psi_{\{y:P_s(y,A)<c\}}(y)P_t(y,\{z\colon P_s(z,A)\geqq c\}) \tag{1.3.1}$$

converges in $P_u(x,\cdot)$-measure to zero as $t\to 0$ as well. This clearly implies that

$$(1.3.2)\qquad \psi_{\{y:P_s(y,A)<c\}}(y)P_t(y,\{z:P_s(z,A)<c\})$$

converges in $P_u(x,\cdot)$-measure to one as $t\to 0$. Now using the first part of Lemma 1.1 with $A=X$, $B=\{y:P_s(y,A)\geqq c\}$, we have that

$$\int_B P_t(y,\{z:P_s(z,A)<c\})P_u(x,dy)$$
$$=P_{t+u}(x,\{z:P_s(z,A)<c\})-\int_{cB}P_t(y,\{z:P_s(z,A)<c\})P_u(x,dy),$$

from which it follows that

$$\psi_{\{y:P_s(y,A)\geqq c\}}(y)P_t(y,\{z:P_s(z,A)<c\})$$

tends to zero in $P_u(x,\cdot)$-measure as $t\to$. The "$>$" part of the lemma follows by a similar argument, and the "$\leqq$" and "$\geqq$" parts from these by taking complements.

LEMMA 1.4. *If $\{A_n\}$ is a sequence of sets such that for each n and for some fixed $u>0$ and $x\in X$, $P_t(y,A_n)$ converges to $\delta_{A_n}(y)$ in $P_u(x,\cdot)$-measure as $t\to 0$, then it follows that $P_t(y,B)$ converges to $\delta_B(y)$ in $P_u(x,\cdot)$-measure as well, where $B=\bigcup_{n=1}^{\infty}A_n$.*

Proof. Note first that, letting $B_k=\bigcup_{n=1}^{k}A_n$,

$$\int_{B_k}\{1-P_t(y,B_k)\}P_u(x,dy)\leqq\sum_{n=1}^{k}\int_{A_n}\{1-P_t(y,B_k)\}P_u(x,dy)$$
$$\leqq\sum_{n=1}^{k}\int_{A_n}\{1-P_t(y,A_n)\}P_u(x,dy),$$

and that

$$\int_{cB_k}P_t(y,B_k)P_u(x,dy)=\sum_{n=1}^{k}\int_{B_k}P_t(y,A_n)P_u(x,dy)$$
$$\leqq\sum_{n=1}^{k}\int_{cA_n}P_t(y,A_n)P_u(x,dy),$$

and thus that the result of the lemma follows for finite sums. Next, to see the result for countably infinite sums, note that for each $k>1$,

$$\int_B\{1-P_t(y,B)\}P_u(x,dy)\leqq\int_B\{1-P_t(y,B_k)\}P_u(x,dy).$$

Hence, using the result for finite sums, we have for each $k>1$,

$$\int_B \{1-P_t(y,B)\}P_u(x,dy) \leqq P_u(x,B_k),$$

and thus that

$$\lim_{t\to 0}\int_B \{1-P_t(y,B)\}P_u(x,dy) = 0. \tag{1.4.1}$$

It remains to show that

$$\lim_{t\to 0}\int_{cB} P_t(y,B)P_u(x,dy) = 0.$$

From the first part of Lemma 1.1, putting $A = X$, and interchanging B with cB, we have that

$$\int_{cB} P_t(y,B)P_u(x,dy) = P_{t+u}(x,B) - \int_B P_t(y,B)p_u(x,dy)$$

and the desired result follows from this, and (1.4.1).

Definition 1.1. Let $\mathcal{S}$ be the Borel field generated by sets of the form $\{x\colon P_t(x,A) \in (a,b]\}$ for each $t > 0$, $A \in \mathcal{B}$ and a and b real.

THEOREM 1.1. *If $G \in \mathcal{S}$, $x \in X$, and $u > 0$, then $P_t(y,G)$ converges to $\delta_G(y)$ in $P_u(x,\cdot)$-measure, as $t \to 0$.*

Proof. Since if $P_t(y,A)$ converges to $\delta_A(y)$ in $P_u(x,\cdot)$-measure as $t\to 0$ then $P_t(y,cA)$ converges to $\delta_{cA}(y)$ in $P_u(x,\cdot)$-measure as $t\to 0$ as well, it follows easily from Lemma 1.4 that the class of sets for which the assertion holds is a monotone class. That it includes finite unions of sets of the form $\{x\colon P_t(x,A)\in(a,b]\}$ follows from Lemmas 1.3 and 1.4.

LEMMA 1.5. *If $\{(X,\mathcal{B},\mu_\gamma),\gamma\in\Gamma\}$ is a family of bounded measure spaces, and if $\{f_n(x)\}$ is a sequence of $\mathcal{B}$-measurable function which, for each $\gamma\in\Gamma$, converges in μ_γ-measure, then there exists a function $f(x)$, measurable with respect to the Borel field $\bigcap_{\gamma\in\Gamma}\mathcal{B}(\mu_\gamma)$, where $\mathcal{B}(\mu_\gamma)$ denotes the completion of $\mathcal{B}$ with respect to μ_γ, such that for each $\gamma\in\Gamma$, $\{f_n(x)\}$ converges in μ_γ-measure to $f(x)$.*

Proof. For each fixed $\gamma\in\Gamma$ there exists a subsequence $\{n^\gamma(k)\}$ and a set A_γ such that

(i) $A_\gamma = \{x\colon \lim_{k\to\infty} f_{n^\gamma(k)}(x) \text{ exists}\}$,

(ii) $\mu_\gamma(A_\gamma) = 1$,

(iii) $f_n(x)$ converges in μ_γ-measure to $\lim_{k\to\infty} f_{n^\gamma(k)}(x)$ on A_γ.

Clearly, the family $\{A_\gamma, \gamma \in \Gamma\}$ of sets of $\mathcal{B}$ has at most continuum distinct elements, since there are at most continuum distinct subsequences $\{n^\gamma(k)\}$, $\gamma \in \Gamma$. For this reason there exists a well-ordering of these sets such that for each $A_{\gamma'}$ there are at most countably many distinct sets $A_\gamma, A_\gamma < A_{\gamma'}$. On

$$A'_{\gamma'} = A_{\gamma'} - \bigcup_{A_\gamma < A_{\gamma'}} A_\gamma,$$

define $f(x)$ by putting $f(x) = \lim_{k\to\infty} f_{n^{\gamma'}(k)}(x)$. On $X - \bigcup_{\gamma\in\Gamma} A_\gamma$ define $f(x)$ by putting $f(x) = 0$. It follows that $f(x)$ is defined everywhere, and by (ii) it follows that $f(x)$ is measurable with respect to $\bigcap_{\gamma\in\Gamma} \mathcal{B}(\mu_\gamma)$.

To see that for each $\gamma' \in \Gamma$, $\{f_n(x)\}$ converges in $\mu_{\gamma'}$-measure to $f(x)$, note that for each $\epsilon > 0$,

$$\begin{aligned}\lim_{n\to\infty} \mu_{\gamma'}\{|f_n - f| > \epsilon\} &= \lim_{n\to\infty} \mu_{\gamma'}\{A_{\gamma'} \cap \{x : |f_n - f| > \epsilon\}\}\\ &= \lim_{n\to\infty} \mu_{\gamma'}\{\bigcup_{A'_\gamma \leqq A_{\gamma'}} A'_\gamma \cap \{x : |f_n - f| > \epsilon\}\}\\ &= \lim_{n\to\infty} \sum_{A'_\gamma \leqq A_{\gamma'}} \mu_{\gamma'}\{A'_\gamma \cap \{x : |f_n - f| > \epsilon\}\}.\end{aligned}$$

Note that there are at most countably many terms in the last sum, and that each term of the sum tends to zero as $n \to 0$. That the whole sum tends to zero follows from the fact that

$$\sum_{A'_\gamma \leqq A_{\gamma'}} \mu_{\gamma'}\{A'_\gamma\} = K,$$

when K is the bound on the measures.

Definition. Let $\mathcal{S}_1 = \bigcap_{\substack{x\in X\\ t\in(0,+\infty)}} \mathcal{S}_{(x,t)}$, where $\mathcal{S}_{(x,t)}$ denotes the completion of $\mathcal{S}$ with respect to $P_t(x, \cdot)$, and let $\mathcal{B}_1$ be the smallest Borel field containing $\mathcal{S}_1$ and $\mathcal{B}$.

THEOREM 1.2. *If for each $A \in \mathcal{B}$, $u > 0$, and $x \in X$ we have that $P_t(y, A)$ converges, as $t \to 0$, in $P_u(x, \cdot)$ measure, then there exist functions $U(y, A)$, defined for all $y \in X$ and $A \in \mathcal{B}_1$ such that:*

(i) $U(\cdot, A)$ *is $\mathcal{S}_1$-measurable, for each $A \in \mathcal{B}$,*

(ii) $U(y, G) = \delta_G(y)$, *for each $G \in \mathcal{S}_1$,*

(iii) $U(y, \bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} U(y, A_i)$

for every sequence $\{A_i\}$ of pairwise disjoint sets of $\mathcal{B}_1$ except on a set N

(depending on $\{A_i\}$) such that $P_t(x,N)=0$ for each $t>0$ and $x\in X$, and

$$\text{(iv)}\quad \lim_{s\to 0}\int |P_s(y,A)-U(y,A)|\,P_t(x,dy)=0,$$

for each $x\in X$, $t>0$, and $A\in\mathcal{B}$.

Proof. Define $U(x,G)=\delta_G(x)$ for each $G\in\mathcal{G}_1$, and choose a version of the limit measure for $U(x,A)$, for $A\in\mathcal{B}_1$ and $A\notin\mathcal{G}_1$, as is possible by Lemma 1.5. (i), (ii), (iii), and (iv) follow by Lemma 1.5, definition, by a neasy proof, and hypothesis, respectively.

COROLLARY 1.1. *Under the hypothesis of Theorem 1.2, if $A\in\mathcal{B}$ and $\epsilon>0$, then there exist positive constants $a_1,a_2,\cdots,a_N$, and sets $A_1,\cdots,A_N$ of $\mathcal{G}$ such that*

$$|P_t(x,A)=\sum_{k=1}^{N}a_kP_t(x,A_k|<\epsilon$$

uniformly in x and t.

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] D. G. Austin, R. M. Blumenthal and R. V. Chacon, "On continuity of transition functions," *Duke Mathematical Journal*, vol. 25 (1958), pp. 533-542.

[2] D. Blackwell, "Idempotent Markoff chains," *Annals of Mathematics*, vol. 43 (1942), pp. 560-567.

[3] J. L. Doob, "Topics in the theory of Markoff chains," *Transactions of the American Mathematical Society*, vol. 52 (1942), pp. 529-540.

COLLINEATION GROUPS OF NON-DESARGUESIAN PLANES II.*

Some seminuclear division algebras.

By D. R. Hughes.[1]

1. Introduction. We will consider a class of non-associative division algebras and the projective planes they coordinatize, with the aim of determining the collineation groups of the planes. This will be the second class of finite division ring planes so analyzed, and we will see that like the "twisted fields" of Albert ([2]), the groups are solvable. This is in sharp contrast to the finite Hall Veblen-Wedderburn planes and the (non-Veblen-Wedderburn) Hughes planes, both of which have non-solvable collienation groups ([4, 6]). Since both the Hall and Hughes planes appear more removed from the Desarguesian case than any division ring plane, it is striking that all the known finite non-associative division ring planes have solvable groups.

Our treatment of the collineation group leaves unanswered a number of possibly interesting questions: (1) Since we only determine a sub-normal series for the group, is the group itself amenable to direct computation? (2) What is the transitive structure of the group, and more particularly, what are the transitive constituents on the line at infinity of the autotopism group $\mathfrak{G}$?

2. Preliminary discussion. Let R be a division ring; i.e., $(R,+)$ is an abelian group, both distributive laws hold, there is a multiplicative identity 1 $(\neq 0)$, and every equation $ax=b$ $(ya=b)$, for $a\neq 0$, has a unique solution for x (for y). Let T be an additive one-to-one mapping of R upon R, and a, b non-zero elements of R; the triple (T,a,b) is called an *autotopism* of R if

$$(1)\qquad (xy)T=(xa)T(by)T, \text{ for all } x, y \text{ in } R.$$

Let us suppose that R is non-alternative, and let π be the projective

* Received February 24, 1959.

[1] This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 18(600)-1383.

plane coordinatized by R (see [4, 5] for construction of π from R). Then there is a metabelian group $\mathfrak{G}_0$, generated by "translations" and "shears," such that if $\mathfrak{G}_1$ is the full collineation group of π, then $\mathfrak{G}_0$ is normal in $\mathfrak{G}_1$ and $\mathfrak{G}_1/\mathfrak{G}_0$ is isomorphic to the group $\mathfrak{G}$ of all autotopisms of R (see [2, 5]). So we will content ourselves with studying $\mathfrak{G}$. We remark that in case R is finite, to say R is non-alternative is equivalent to saying that it is not a field.

Now we introduce some notation. The *right nucleus* N_r of R is the set of all n in R such that $(xy)n = x(yn)$, for all x, y in R; the *middle* and *left* nuclei are defined analogously. The *nucleus* of R is the intersection of the three one-sided nuclei, while the *center* of R is the set of all z in the nucleus such that $xz = zx$ for all x in R. It is well-known that all five of these subsets are themselves division rings (even associative, of course).

If x is in R, define the mappings R_x and L_x of R by $yR_x = yx$, $yL_x = xy$. If $x \neq 0$, these mappings are non-singular and have inverses. Sometimes we will write $R(x)$ for R_x, etc.

Now let (T, a, b) be an autotopism of R, and for each n in N_r, define $n^\alpha = (ban)T$. Then (see [3, p. 250]), α is an automorphism of N_r onto N_r; let x be in R, n in N_r. Then:

$$(xn)T = [(xR_a^{-1} \cdot a)n]T = [(xR_a^{-1})(an)]T,$$

and so, from (1):

$$(xn)T = [(xR_a^{-1})a]T(ban)T = (xT)n^\alpha.$$

LEMMA 1. *Every autotopism of R is a semi-linear transformation over the right nucleus.*[2]

Now we define our class of division rings. Let F be a field, σ a non-identity automorphism of F, and δ_0 and δ_1 elements of F such that

$$\delta_0 \neq w^{1+\sigma} + \delta_1 w, \text{ for any } w \text{ in } F. \tag{2}$$

Note that (2) requires $\delta_0 \neq 0$. Furthermore, if $\sigma^2 = 1$, $\delta_1 = 0$, and F is finite, then $\delta_0 \neq \delta_0^\sigma$. For the elements $x^{1+\sigma}$, as x ranges over F, range over the subfield K of all elements fixed by σ, and so δ_0 cannot be in this subfield. Let R be a two-dimensional vector space over F, with basis elements 1, λ. Define multiplication in R by

$$(x + \lambda y)(u + \lambda v) = (xu + \delta_0 y^\sigma v) + \lambda(yu + x^\sigma v + \delta_1 y^\sigma v). \tag{3}$$

This corresponds to demanding that $\lambda^2 = \delta_0 + \lambda\delta_1$, $x\lambda = \lambda x^\sigma$, and that F be

[2] There is an analogous theorem for the left nucleus, and even for the center.

the right and middle nuclei of R. Then R is a division ring[3]; for the rest of the paper, we shall assume that R is finite.

We wish to consider the space $\mathcal{S}$ of all additive operators on F, and in particular that subspace of $\mathcal{S}$ containing all the mappings R_x, x in F, as well as the mappings $Q: x \rightarrow x^{\mu}$, where μ is an automorphism of F. We will further consider the set of 2 by 2 matrices over $\mathcal{S}$, considered as acting on elements of R, where $x + \lambda y$ is identified with the vector (x, y).

Let σ be the defining automorphism of R, and S the element of $\mathcal{S}$ defined by $S: x \rightarrow x^{\sigma}$; for an autotopism (T, a, b) of R, let α be the automorphism of the right nucleus of R, as in Lemma 1 and the preceding discussion. Since $F = N_r$, α is an automorphism of F; define $A: x \rightarrow x^{\alpha}$. For simplicity, write $\gamma = \alpha^{-1}$.

Lemma 1 asserts that T has the form AT_1, where

$$T_1 = \begin{bmatrix} R(a_0) & R(a_1) \\ R(b_0) & R(b_1) \end{bmatrix}.$$

Furthermore, from (3) we can write:

$$R_{u+\lambda v} = R(u)I + SR(v)E,$$

where the terms on the right are understood to be among the two-dimensional matrices over $\mathcal{S}$, and the term on the left is the ordinary right multiplication in R. Here, E and I are defined by:

$$I = \begin{bmatrix} R(1) & 0 \\ 0 & R(1) \end{bmatrix}, \qquad E = \begin{bmatrix} 0 & R(1) \\ R(\delta_0) & R(\delta_1) \end{bmatrix}.$$

Now we note that:

$$AR(x) = R(x^{\gamma})A \tag{4}$$

$$R(x)S = SR(x^{\sigma}). \tag{5}$$

Furthermore, S and A commute with each other, since F is a Galois field.

Let the a of (T, a, b) be $a = r + \lambda s$, and write $yP = (by)T$; then (1) is equivalent to:

$$R_{r+\lambda s}TR_{yP} = R_yT. \tag{6}$$

Let $Q = P^{-1}$; then (6) becomes:

$$R_{r+\lambda s}TR_{u+\lambda v}T^{-1} = R_{(u+\lambda v)Q}, \tag{7}$$

[3] In a forthcoming paper, these division rings are studied in more detail by Edwin Kleinfeld and the present writer.

or:

$$(8)\qquad [R(r)I + SR(s)E]AT_1[R(u)I + SR(v)E]T_1^{-1}A^{-1} = R_{(u+\lambda v)Q}.$$

3. Computation of the autotopism group. If we multiply out the left side of (8) and use (4) and (5), we find:

$$(9)\qquad R(ru^\gamma)I + S[R(r^\sigma v^\gamma)T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} + R(su^\gamma)E] + S^2R(s^\sigma v^\gamma)E^\sigma T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} = R_{(u+\lambda v)Q}.$$

Here T_1^γ, for instance, means:

$$T_1^\gamma = \begin{bmatrix} R(a_0^\gamma) & R(a_1^\gamma) \\ R(b_0^\gamma) & R(b_1^\gamma) \end{bmatrix}.$$

Since the left side of (9) must be a right multiplication, it must have the form $R(x)I + SR(y)E$, for some x, y in F. Now we must distinguish cases.

If $S^2 \neq 1$ (i.e., if $\sigma^2 \neq 1$), then the last term on the left of (9) can only be identically zero; this implies $s = 0$. Then (9) becomes:

$$(10)\qquad R(ru^\gamma)I + SR(r^\sigma v^\gamma)T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} = R_{(u+\lambda v)Q}.$$

So the second term in (10) has the form $SR(y)E$, and hence:

$$(11)\qquad T_1^{\sigma\gamma}E^\gamma = R(y)ET_1^\gamma, \text{ for some } y \neq 0.$$

On the other hand, if $\sigma^2 = 1$, then $S^2 = 1$, so (9) becomes:

$$(12)\qquad R(ru^\gamma)I + R(s^\sigma v^\gamma)E^\sigma T_1^{\sigma\gamma}T_1^{-\gamma} + S[R(r^\sigma v^\gamma)T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} + R(su^\gamma)E] = R_{(u+\lambda v)Q}.$$

In the following, we can simplify matters by understanding that:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad E = \begin{bmatrix} 0 & 1 \\ \delta_0 & \delta_1 \end{bmatrix}, \qquad T_1 = \begin{bmatrix} a_0 & a_1 \\ b_0 & b_1 \end{bmatrix}.$$

Then (12) implies:

$$(13)\qquad ru^\gamma I + s^\sigma v^\gamma E^\sigma T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} = kI,$$

$$(14)\qquad r^\sigma v^\gamma T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} + su^\gamma E = mE,$$

for some choice of k, m in F. So we write, from (13) and (14):

$$(15)\qquad s^\sigma v^\gamma E^\sigma T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} = (k - ru^\gamma)I,$$

$$(16)\qquad r^\sigma v^\gamma T_1^{\sigma\gamma}E^\gamma T_1^{-\gamma} = (m - su^\gamma)E.$$

Then (15) and (16) imply:

$$s^\sigma(m - su^\gamma)E^\sigma E = r^\sigma(k - ru^\gamma)I. \tag{17}$$

Now note that

$$E^\sigma E = \begin{bmatrix} \delta_0 & \delta \\ \delta_0\delta_1^\sigma & \delta_0^\sigma + \delta_1^{1+\sigma} \end{bmatrix}.$$

So the left side of (17) is not scalar unless it is 0 or $\delta_1 = 0$ and $\delta_0 = \delta_0^\sigma$; as remarked earlier, this latter possibility violates (2) since $\sigma^2 = 1$. But the right side of (17) is scalar, so either one of r, s is zero, or $m = su^\gamma$ and $k = ru^\gamma$. Suppose the latter of these two possibilities occurs; then $(u + \lambda v)Q = k + \lambda m = (r + \lambda s)u^\gamma$, and this implies that Q is singular, which is impossible. Hence:

LEMMA 2. *If $\sigma^2 = 1$, then $r = 0$ or $s = 0$.*

If $s = 0$ then it is easily seen that we are led to equation (11), and so no distinction is necessary between $\sigma^2 = 1$ and $\sigma^2 \neq 1$. On the other hand, if $r = 0$, the additional equation necessary is:

$$E^\sigma T_1^\sigma E^\gamma = nT_1^\gamma, \text{ for some } n \neq 0. \tag{18}$$

Now suppose we have two autotopisms A_1T_1 and A_2T_2 (strictly, we are looking only at the additive map part of an autotopism triple). Then one easily computes that $(A_1T_1)(A_2T_2) = (A_1A_2)(T_1^\alpha T_2)$, where α is the automorphism of N_r associated with the autotopism A_2T_2. Remembering that the γ in (18) is actually α^{-1}, it is straightforward to show:

LEMMA 3. *The product of two autotopisms satisfying* (18) *is an autotopism with $s = 0$ and satisfying* (11).

So if we disregard the autotopisms satisfying (18), we will be considering a subgroup $\bar{\mathfrak{G}}$ of $\mathfrak{G}$, where $\mathfrak{G}/\bar{\mathfrak{G}}$ has order one or two. We shall restrict attention to this group $\bar{\mathfrak{G}}$, and thus to autotopisms satisfying:

$$T^{\sigma\gamma}E^\gamma = kET^\gamma, \tag{19}$$

where k is in F, $k \neq 0$, and E and T are the matrices E and T_1 following equation (12).

Among all solutions of (19), the set $\mathfrak{H}_1$ of those T for which $\gamma = 1$ is a normal subgroup, whose factor group is contained in the group of all semilinear transformations over F modulo the group of linear transformation over F. I. e., $\bar{\mathfrak{G}}/\mathfrak{H}_1$ is isomorphic to a subgroup of the automorphism group of F; since F is a Galois field, $\bar{\mathfrak{G}}/\mathfrak{H}_1$ is cyclic.

So $\mathfrak{H}_1$ consists of the T satisfying:

$$T^{\sigma}E = kET. \tag{20}$$

But among all elements of $\mathfrak{H}_1$, the subset $\mathfrak{H}$ of all solutions of (20) for which $k = 1$ is a normal subgroup (of $\mathfrak{H}_1$), and $\mathfrak{H}_1/\mathfrak{H}$ is isomorphic to a subgroup of the multiplicative group of F. (This can be easily seen by considering the mapping which sends a solution of (20) onto the element k associated with it; the mapping is a homomorphism.)

Now we have reduced our problem to the point where some direct computing is possible. $\mathfrak{H}$ consists of all (non-singular) 2 by 2 matrices T over F which satisfy:

$$T^{\sigma}E = ET. \tag{21}$$

Then (21) immediately implies:

$$b_0 = a_1^{\sigma}\delta_0, \tag{22}$$

$$b_1 = a_0^{\sigma} + a_1^{\sigma}\delta_1, \tag{23}$$

$$a_0^{\sigma^2} + a_1^{\sigma^2}\delta_1^{\sigma} = a_0 + a_1^{\sigma}\delta_1, \tag{24}$$

$$a_1^{\sigma^2}\delta_0^{\sigma} + a_0^{\sigma^2}\delta_1 + a_1^{\sigma^2}\delta_1^{1+\sigma} = a_1\delta_0 + a_0^{\sigma}\delta_1 + a_1^{\sigma}\delta_1^{2} \tag{25}$$

So if T is in $\mathfrak{H}$, we can represent T by the vector (a_0, a_1), consisting of its first row, where a_0, a_1 satisfy (24) and (25). The multiplication of these vectors (i.e., the multiplication of the elements of $\mathfrak{H}$) is:

$$(a_0, a_1)(c_0, c_1) = (a_0c_0 + \delta_0 a_1 c_1^{\sigma}, a_0c_1 + a_1c_0^{\sigma} + \delta_1 a_1 c_1^{\sigma}), \tag{26}$$

as one sees by multiplying together two matrices satisfying (21) and using (22) and (23).

Now consider the division ring R' which is anti-isomorphic to R; we can represent the multiplication in R' by:

$$(x + y\lambda)(u + v\lambda) = (xu + \delta_0 y v^{\sigma}) + (xv + yu^{\sigma} + \delta_1 y v^{\sigma})\lambda.$$

But this argrees with (26) if we identify $(a_0\, a_1)$ with $a_0 + a_1\lambda$, and so $\mathfrak{H}$ is isomorphic to a subgroup of the multiplicative loop of R'. In order to identify $\mathfrak{H}$ more precisely, we find it convenient to compute the right nucleus of R'.

For any three elements a, b, c of R', let $(a, b, c) = (ab)c - a(bc)$. Then it is straightforward computation that:

$$\begin{aligned}(x + y\lambda, u + v\lambda, w + z\lambda) = {} & yv^{\sigma}\delta_0[w - w^{\sigma^2} + \delta_1 z^{\sigma} - \delta_1 z^{\sigma^2}] \\ & + yv^{\sigma}[\delta_0 z - \delta_0^{\sigma} z^{\sigma^2} + \delta_1(\delta_1 z^{\sigma} - \delta_1^{\sigma} z^{\sigma^2}) + \delta_1(w^{\sigma} - w^{\sigma^2})]\lambda.\end{aligned}$$

So $w + z\lambda$ is in the right nucleus of R' if and only if:

$$w + \delta_1 z^\sigma = w^{\sigma^2} + \delta_1^\sigma z^{\sigma^3}, \tag{27}$$

$$\delta_0 z + \delta_1^2 z^\sigma + \delta_1 w^\sigma = \delta_0^\sigma z^{\sigma^2} + \delta_1^{1+\sigma} z^{\sigma^2} + \delta_1 w^{\sigma^2}. \tag{28}$$

But (27), (28) become equations (24), (25) with the change $a_0 = w$, $a_1 = z$, and thus $\mathfrak{H}$ is isomorphic to the multiplicative group of the right nucleus of R'. Anti-isomorphic groups are isomorphic, so this means that $\mathfrak{H}$ is isomorphic to the multiplicative group of the left nucleus of R; the left nucleus is finite and associative, hence is a Galois field, so it has a cyclic multiplicative group.

THEORM 1. *The autotopism group $\mathfrak{G}$ of R is solvable when R is finite, and $\mathfrak{G}$ has a sub-normal series:*

$$\mathfrak{G} \supseteq \bar{\mathfrak{G}} \supseteq \mathfrak{H}_1 \supseteq \mathfrak{H} \supseteq 1,$$

where (i) *$\mathfrak{G}/\bar{\mathfrak{G}}$ has order one or two, and has order one if $\sigma^2 \neq 1$;* (ii) *$\bar{\mathfrak{G}}/\mathfrak{H}_1$ is isomorphic to a subgroup of the automorphism group of F;* (iii) *$\mathfrak{H}_1/\mathfrak{H}$ is isomorphic to a subgroup of the multiplicative group of F;* (iv) *$\mathfrak{H}$ is isomorphic to the multiplicative group of the left nucleus of R.*

COROLLARY. *If π is the projective plane coordinatized by the finite semi-nuclear division algebra R, then the collineation group of π is solvable.*

THE UNIVERSITY OF CHICAGO.

REFERENCES.

[1] A. A. Albert, "On nonassociative division algebras," *Transactions of the American Mathematical Society*, vol. 72 (1952), pp. 296-309.

[2] ———, "On the collineation groups associated with twisted fields," *Bulletin of the Calcutta Mathematical Society*, to appear.

[3] R. H. Bruck, "Contributions to the theory of loops," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 245-354.

[4] D. R. Hughes, "Collineation groups of non-Desarguesian planes I. The Hall Veblen-Wedderburn systems," *American Journal of Mathematics*, vol. 81 (1959), pp. 921-938.

[5] Günter Pickert, *Projektive Ebenen*, Berlin, 1955.

[6] Guido Zappa, "Sui gruppi di collineazioni dei piani di Hughes," *Bollettino della Unione Matematica Italiana*, vol. 12 (1957), pp. 507-516.

TAME COVERINGS AND FUNDAMENTAL GROUPS OF ALGEBRAIC VARIETIES.*

Part II: Branch Curves with Higher Singularities.

By Shreeram Abhyankar.[1]

TABLE OF CONTENTS

* Received May 15, 1959.

[1] Alfred P. Sloan Foundation Research Fellow.

Introduction. In this paper we study tame coverings and fundamental groups of a non-singular algebraic surface V (over an algebraically closed ground field k) minus a curve W having arbitrary singularities; it is a continuation of Part I of this series [*American Journal of Mathematics*, vol. 81, 1959] in which we considered the situation when W had only strong normal crossings (there the dimension of V was arbitrary and W was of co-dimension one). Here, in the introduction, we shall briefly and approximately describe the main results and the contents of the various sections, referring for the precise definitions and statements to the body of this paper and to Part I. Denote by $\pi'(V-W)$ the group tower of the galois groups over $k(V)$ of all the finite galois extensions of $k(V)$ (in some fixed algebraic closure of $k(V)$), which are tamely ramified over V and for which the branch locus over V is contained in W. For an irreducible component W_1 of W define the strength of singularities $\nu(W_1, W; V)$ by $\nu(W_1, W; V) = \frac{1}{2}\sum \mu_i(\mu_i + 1)$, where the μ_i are the multiplicities of the various points and "infinitely near" points of W_1 at which W does not have a strong normal crossing. Let $W_1, \cdots, W_t$ be the irreducible components of W. Then the main results on fundamental groups in this paper are: (1) If V is simply connected and for some labelling of the components W_j we have $\dim |W_j| > 1 + \nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; V)$ for $j = 1, \cdots, t$, then $\pi'(V-W)$ is generated by t generators, is t-step solvable, and has a weak parent group generated by t generators. (2) If V is simply connected and $\dim |W_j| > 1 + \nu(W_j, W; V)$ for $j = 1, \cdots, t$, then $\pi'(V-W)$ is generated by t generators, is t-step nilpotent, and has a weak parent group generated by t generators; if in addition W_j and W_k have a point in common at which W has a normal crossing whenever $j \neq k$, then $\pi'(V-W)$ is abelian.

An analysis of singularities only for curves on an algebraic surface would have been quicker, and would have been adequate for the above results on fundamental groups; however, we thought it appropriate here to develop systematically an analysis of singularities in an arbitrary two dimensional regular local domain, and also to include some related considerations for local domains; this accounts for the length of the paper. Presently we shall describe the contents of the various sections.

Section 1. Notations and conventions are fixed and some preliminary remarks are made.

Section 2. Some auxiliary lemmas, mainly on local rings, are proved. This section need not be read in the beginning; and the reader may look up the relevant parts of it when referred to in the following sections.

Section 3. Here is introduced the notion of an m-th quadratic transform (for any non-negative integer m) of a regular local domain R of dimension greater than one, a quadratic transform of R is then an m-th quadratic transform of R for some m. It is proved that there is a natural one to one correspondence between the set of all quadratic transforms of R and the set of all quadratic transforms of the completion $\bar{R}$ or R; in this correspondence an m-th quadratic transform of R corresponds to an m-th quadratic transform of $\bar{R}$ (Proposition 1).

Section 4. Let R be a two dimensional regular local domain, let M be the maximal ideal in R, let A and B be non-zero principal ideals in R, and let S be a two dimensional regular local domain having the same quotient field as R and having center M in R. The R-leading degree of A is denoted by $\lambda_R(A)$ and is called the multiplicity of A at R. The notion of the S^R transform of A is introduced, it is denoted by $S^R[A]$ and its S-leading degree is denoted by $\mu_{S,R}(A)$ (Definition 3). There is a natural one to one correspondence between the immediate (i.e., first) quadratic transforms of R and irreducible forms in two variables over R/M (Lemma 13). The notion of a valuation branch of A at R is introduced, the set of all valuation branches of A at R is a finite set and it is denoted by $\mathfrak{S}(A,R)$ (Definition 4), there is a one to one natural correspondence between the valuation branches of A at R and the analytic branches of A at R (Lemma 14). If A and B are co-prime, then so are $S^R[A]$ and $S^R[B]$, if A is primary then so is $S^R[A]$, the operation of taking transforms is transitive in S and is multiplicative in the primary factors of A, if $\lambda_R(A) \leqq 1$ then $\mu_{S,R}(A) \leqq 1$,[2] and finally $\mathfrak{S}(A,R) \supset \mathfrak{S}(S^R[A],S)$ (Lemma 15). For each m, there are at most a finite number (and at least one if $A \neq S$) of m-th quadratic transforms R_m of R through which A passes (i.e., $R_m{}^R[A] \neq R_m$) (Lemma 16). The concept of a quadratic transform of an algebraic surface is recalled (Definition 5). If R is complete and z is a non-zero irreducible non-unit in R then the reduced R-leading form of z cannot have coprime factors in $(R/M)[X,Y]$ (Lemma 17), this is false for dimension > 2 (Remark 1).

Section 5. Let R be a regular local domain and let A be a principal ideal in R. The notion of A to have a normal crossing at R, and the notion of A to have a strong normal crossing at R are introduced (Definition 6). Let S be a quadratic transform of R. If A has a normal crossing (respectively: strong normal crossing) at R then $S^R[A]$ has a normal crossing

[2] More generally, it can be shown that always $\mu_{S,R}(A) \leqq \lambda_R(A)$.

(respectively: a strong normal crossing) at S (Lemma 18). If R is the quotient ring of a point on an algebraic surface, if A has a normal crossing at R, and if $S \neq R$, then $S^R[A]$ has a strong normal crossing at S (Lemma 19); this is false for dimension > 2 (Remark 2). If R is two dimensional and is either algebraic or absolute then the singularities of A can be resolved by applying quadratic transformations to R (Proposition 2).

Section 6. Let A and B be non-zero principal ideals in a two dimensional regular local domain R. The strength of singularity of A on B at R is introduced and is denoted by $\nu(A, B; R)$ (Definitions 7, 8); and via it, the notion of the strength of singularities of a curve W on another curve W^* on a non-singular algebraic surface V is introduced by taking the sum of the strengths of singularities at all the points of V, it is denoted by $\nu(W, W^*; V)$ (Definition 9). $\nu(A, B; R)$ is an analytic invariant and hence can be defined by using completions only (Proposition 3); $\nu(A, B; R)$ is finite if A is a product of distinct prime ideals and R is algebraic or absolute (Proposition 4); consequently $\nu(W, W^*; V)$ is finite (Definition 9). If W is an irreducible component of W^*, if $\dim |W| > 1 + \nu(W, W^*; V)$, then there exists a quadratic transform (V^*, f) of V such that $f^{-1}(W^*)$ has only strong normal crossings on $f^{-1}[W]$ and $\dim |f^{-1}[W]| > 1$ (Proposition 5); this is proved by applying successive immediate quadratic transforms to V and estimating at each stage the decrease in $\dim |W|$.

Section 7. The strength of a singularity is computed for the following cases: contact of two simple branches (of arbitrary order), ordinary point (several simple branches with distinct tangents), cusps (of arbitrary order), composite cusps with distinct tangents, etc. All these notions are developed and the corresponding computations are made in the set up of an arbitrary two dimensional regular local domain. The results of this section are used only in Section 10; hence the reading of this section may be postponed until then.

Section 8. For dimension two, a direct proof of the abelian character of local galois groups over a branch curve having a normal crossing is given (Proposition 11β) without using "Purity of branch locus"; and "Purity" is derived from it as a corollary. Most of the techniques used in the proof carry over for arbitrary two dimensional regular local domains and hence the proof can probably be geenralized to this general case (Remark 4). Above a strong normal crossing of the apparent branch locus, there can be no local splitting (Proposition 12β); this is in general false for normal crossings which

are not strong normal crossings (Remark 5); this necessitates the replacement of "normal crossing" by "strong normal crossing" in some of the results of Part I. This, together with other minor corrections to Part I, is given in Remark 6.

Section 9. Using Proposition 5 of Section 6, and then applying the technique of the proof of Proposition 6 of Section 11 of Part I, the following refinement of the quoted result of Part I is obtained: If W is a curve on a non-singular algebraic surface and W_1 is an irreducible component of W such that $\dim |W_1| > 1 + \nu(W_1, W; V)$, if V^* is a tamely ramified covering of V and ϕ is the rational map of V^* onto V, and if the branch locus of V^* over V is contained in W, then $\phi^{-1}(W_1)$ is irreducible (Proposition 14). In view of this refinement of Proposition 6 of Part I, the main results of this section now follows by essentially carrying over the proofs of the corresponding results of Sections 11 and 12 of Part I. For dimension two, the main results of Part I are now subsumed under the results of this section.

Section 10. For dimension two, the results of Section 9 now give the corresponding refinements of the results of Sections 13, 14, 15 of Part I (including the Theorems of Zariski and Picard) and the latter are now subsumed under the results of this section (Propositions 17, 18, 19 and Theorems 3, 4, 5). Several explicit corollaries of the results of Sections 9 and 10 can be obtained using the computations of Sections 7, some examples of this are given (Examples 1, 2, 3).

1. Conventions and notations. Part I of this series with the subtitle "Branch loci with normal crossings; Applications: Theorems of Zariski and Picard" (which appeared in the *American Journal of Mathematics*, vol. 81 (1959), pp. 46-94.) will be referred to as "Part I." Besides the conventions and notations introduced in Part I we shall use the following additional ones.

In a ring A, the product over an empty set of elements will be one, thus for $a_1, \cdots, a_m$ in A: $a_1 a_2 \cdots a_m = 1$ if $m = 0$; the sum over an empty set of elements will be zero, thus for $a_1, \cdots, a_m$ in A: $a_1 + a_2 + \cdots + a_m = 0$ if $m = 0$; the product or intersection over an empty set of ideals in A will be A itself, thus for ideals $B_1, \cdots, B_m$: $B_1 B_2 \cdots B_m = B_1 \cap B_2 \cap \cdots \cap B_m = A$ if $m = 0$. For an ideal B in a ring A, by $\mathrm{Rad}_A B$ we shall denote the set of all elements a in A such that a^n is in B for some positive integer n; the subscript A may be dropped when it is clear from the context.

For a valuation v of a field K we shall denote by R_v and M_v, the valuation ring of v and the maximal ideal in the valuation ring of v respectively. Let

A be a ring and P an ideal in A; if for a local ring (R, M) we have $R \supset A$ and $M \cap A = P$ then we shall say that R has center P in A; if for a valuation v we have $R_v \supset A$ and $M_v \cap A = P$ then we shall say that v has center P in A. If (R, M) is a local domain and v is a valuation having center M in R, then the transcendence degree of R_v/M_v over R/M will be called the R-dimension of v. If A is a ring and P is a finitely generated ideal in A, and v is a valuation with $R_v \supset R$, then by $v(P)$ we shall denote the minimum of $v(a)$ for a in P; note that if Q is a set of generators of P then $v(P)$ is the minimum of $v(a)$ for a in Q.

A local ring (R, M) is said to be regular, if and only if, R is noetherian, and M has a basis of n elements, where n is the dimension of R. Note that a regular local domain R of dimension two is a unique factorization domain, and hence in R every pure one dimensional ideal is principal, products and intersections of pure one dimensional ideals are the same things, etc.

Now let (R, M) be a regular local domain. For a non-zero ideal A in R there exists a unique integer such that $A \subset M^n$ and $A \not\subset M^{n+1}$, *we shall call n the R-leading degree of A or the multiplicity of A at R* and *we shall denote it by $\lambda_R(A)$*; similarly, for a non zero element a of R the unique integer n, for which $a \in M^n$ and $a \notin M^{n+1}$, will be called the *R-leading degree of a or the multiplicity of a at R and will be denoted* by $\lambda_R(a)$; note that $\lambda_R(a) = \lambda_R(aR)$. Let Q be a representative set of R/M in R (i. e., a subset of R which is mapped one to one onto R/M under the canonical homomorphism of R onto R/M), and let $x = (x_1, \cdots, x_n)$ be a minimal basis of M. Then, for $0 \neq a \in R$ there exists a unique form $f(X_1, \cdots, X_n)$ of degree $\lambda_R(a)$ in the indeterminates $X_1, \cdots, X_n$ with coefficients in Q such that $a - f(x_1, \cdots, x_n)$ is in $M^{\lambda_R(a)+1}$ and we shall call $f(X_1, \cdots, X_n)$ the *R-leading form of a with respect to (Q, x)* and we shall denote it by $\Lambda_{R,Q,x}(a)$; furthermore, the form obtained from $f(X_1, \cdots, X_n)$ by reducing its coefficients modulo M will be called *the reduced R-leading form* of a with respect to x and will be denoted by $\bar{\Lambda}_{R,x}(a)$; for $a = 0$ we set $\Lambda_{R,Q,x}(a) = \bar{\Lambda}_{R,x}(a) = 0$. Again, one or more subscripts of λ, Λ, $\bar{\Lambda}$ may be dropped when they are clear from the context. Note that for $a_1, \cdots, a_m$ in R we have $\bar{\Lambda}(a_1 \cdots a_m) = \bar{\Lambda}(a_1) \cdots \bar{\Lambda}(a_m)$. Also note that, if $y = (y_1, \cdots, y_n)$ is any other minimal basis of M and if $f(X_1, \cdots, X_n) = \bar{\Lambda}_x(a)$ and $g(X_1, \cdots, X_n) = \bar{\Lambda}_y(a)$, then g is obtained from f by applying a non-singular homogeneous linear transformation to $X_1, \cdots, X_n$ with coefficients in R/M; and hence if $f = f_1^{u_1} \cdots f_s^{u_s}$ and $g = g_1^{v_1} \cdots g_t^{v_t}$ are factorizations of f and g where $f_1, \cdots, f_s$ are pair-wise co-prime irreducible forms and $g_1, \cdots, g_t$ are pair-wise co-prime irreducible forms and $u_i, v_i > 0$, then

$s = t$ and after a suitable relabelling $u_i = v_i$, and f_i and g_i are non-zero constant multiples of each other for $i = 1, \cdots, s$.

Now let $x_2, \cdots, x_n$ be elements of R, such that there exists x_1 in R such that $x_1, \cdots, x_n$ is a basis of M; if for the reduced R-leading form $f(X_1, \cdots, X_n)$ of a with respect to $x_1, \cdots, x_n$ we have $f(1, 0, 0, \cdots, 0) \neq 0$, then we shall say that "the line $x_2 = \cdots = x_n = 0$ is non-tangential to a"; note that this does not depend on x_1; also note that if $a_1, \cdots, a_t$ are elements of R, then the line $x_2 = \cdots = x_n = 0$ is non-tangential to $a_1 \cdots a_n$, if and only if, it is non-tangential to each a_i.

Next let V be an n-dimensional irreducible projective algebraic variety over an algebraically closed ground field k. If W is an irreducible subvariety of V, then $P \in W$ if and only if $Q(W, V) \supset Q(P, V)$, and in that case we set $M(P, W, V) = M(W, V) \cap Q(P, V)$; if $P \notin W$, then we set $M(P, W, V) = Q(P, V)$. If W is any subvariety of V with irreducible components $W_1, \cdots, W_t$, then we shall call $M(P, W_1, V) \cap \cdots \cap M(P, W_t, V)$ *the ideal of W at P on V and we shall denote it by* $M(P, W, V)$; note that $P \in W$ if and only if $M(P, W, V) \subset M(P, V)$; *an ideal A in $Q(P, V)$ will be called a defining ideal of W at P on V* if $A = A_1 \cap \cdots \cap A_t$ where A_i is an ideal in $Q(P, V)$ which is primary for $M(P, W_i, V)$; note that then $A_1, \cdots, A_t$ are uniquely determined by A; now assume that P is a simple point of V and that $W_1, \cdots, W_t$ are all of co-dimension 1, then: (1) a defining ideal of W at P on V is exactly a principal ideal A in $Q(P, V)$ such that $\operatorname{Rad} A = M(P, W, V)$; (2) a given ideal A in $Q(P, V)$ is a defining ideal of a pure one co-dimensional subvariety of V if and only if A is a non-zero principal ideal; and (3) a given ideal A in $Q(P, V)$ is *the* ideal of a pure one co-dimensional subvariety of V if and only if A is a non-zero principal ideal which is not contained in the square of any non-unit principal ideal.

If V^* and V are irreducible normal projective algebraic varieties over an algebraically closed ground field k and f is a bi-rational map of V^* onto V which is regular on V^*, and if W is a subvariety of V, then by $f^{-1}[W]$ we shall denote the transform of W under f^{-1} (for definition see [Zariski 15]), and by $f^{-1}(W)$ we shall denote as usual the set theoretic inverse image of the set of points of W. It can immediately be verified that if W is irreducible and of co-dimension one so that $Q(W, V)$ is the valuation ring of a real discrete valuation v of $k(V)/k$, then $f^{-1}[W]$ is also irreducible of co-dimension one and it is the center of v on V^*.

2. Auxiliary lemmas.

LEMMA 1. *Let (R, M) be a regular local domain of dimension $n > 1$ with quotient field K. Let $x_1, \cdots, x_n$ be a minimal basis of M. Let $y_1 = x_1$, $y_i = x_i/x_1$ for $i = 2, \cdots, n$ and let $S = R[y_2, \cdots, y_n]$. Then we have $(y_1 S) \cap R = M$ so that R/M can be canonically identified with a subfield of $S/(y_1 S)$ and then the residues $\bar{y}_2, \cdots, \bar{y}_n$ respectively of $y_2, \cdots, y_n$ modulo $y_1 S$ are algebraically independent over R/M and hence $S/(y_1 S)$ can be identified with the polynomial ring $(R/M)[\bar{y}_2, \cdots, \bar{y}_n]$ in $n-1$ variables. Then the canonical homomorphism of S onto $S/(y_1 S)$ sets up a one to one correspondence between the maximal ideals in S containing y_1 and the maximal ideals in the polynomial ring $S/(y_1 S)$, and hence if P is a maximal ideal in S containing y_1 and $R_1 = S_P \subset K$ and $M_1 = PR_1$ then (R_1, M_1) is a regular local domain of dimension n having center M in R, y_1 is part of a minimal basis of M_1, and R_1/M_1 is a finite algebraic extension of R/M.*

Proof. Follows from [Abhyankar 5, Lemma 3.19 of Section 14].

LEMMA 2. *Let (R, M) and $(\bar{R}, \bar{M})$ be two regular local domains of the same dimension n with quotient fields K and $\bar{K}$ respectively. Assume that K is a subfield of $\bar{K}$, $\bar{R}$ has center M in R and $M\bar{R} = \bar{M}$. Then R/M can be canonically identified with a subfield of $\bar{R}/\bar{M}$; assume that under this indentification $\bar{R}/\bar{M} = R/M$, i.e., given x in $\bar{R}$ there exists y in R such that $x - y$ is in $\bar{M}$. Then given x in R and a positive integer q, there exists y in R such that $x - y$ is in $\bar{M}^q$. Also we have that $K \cap \bar{R} = R$, $K \cap \bar{M} = M$, $M^q\bar{R} = \bar{M}^q$ for any positive integer q, and if A is any ideal in R then $(A\bar{R}) \cap R = A$, and A is principal if and only if $A\bar{R}$ is principal, in particular $\bar{M}^q \cap R = M^q$ for any positive integer q. Consequently, any completion of $\bar{R}$ is also a completion of R.*

Proof. From $M\bar{R} = \bar{M}$ we at once have $M^q\bar{R} = \bar{M}^q$ for all $q > 0$. By induction on q we shall show that given x in $\bar{R}$ there exists y in R such that $x - y$ is in $\bar{M}^q$; for $q = 1$ this is given, so assume that $q > 1$ and that this is true for $q - 1$. Then given x in $\bar{R}$ there exists z in R such that $x - z$ is in $\bar{M}^{q-1}$; since $\bar{M}^{q-1} = M^{q-1}\bar{R}$, $x - z = \sum a_i b_i$ with a_i in M^{q-1} and b_i in $\bar{R}$, the assumption for $q = 1$ tells us that there exists c_i in R and d_i in $\bar{M}$ such that $b_i = c_i + d_i$, let $y = z + \sum a_i c_i$, then $x - y = \sum a_i d_i$ which is in $\bar{M}^q$.

Now let $x_1, \cdots, x_n$ be a minal basis of M. Since $M\bar{R} = \bar{M}$ and dim $\bar{R} = n$; $x_1, \cdots, x_n$ is also a minimal basis of $\bar{M}$. For any non-negative integer q, $y \in M^q$ and $y \notin M^{q+1}$ imply that $y = f(x_1, \cdots, x_n)$ where $f(X_1, \cdots, X_n)$ is a form of degree q with coefficients in R but not all in M; suppose if possible

that $y \in \bar{M}^{q+1}$, then $y = g(x_1, \cdots, x_n)$ where $g(X_1, \cdots, X_n)$ is a form of degree $q+1$ with coefficients in $\bar{R}$, so that $y = h(x_1, \cdots, x_n)$ where $h(X_1, \cdots, X_n)$ is a form of degree q with coefficients in $\bar{M}$; let $t(X_1, \cdots, X_n) = f(X_1, \cdots, X_n) - h(X_1, \cdots, X_n)$; since f has coefficients in R but not all in M and since $\bar{M} \cap R = M$, these coefficients are in $\bar{R}$ but not all in $\bar{M}$; since the coefficients of h are all in $\bar{M}$ we conclude that $t(X_1, \cdots, X_n)$ is a form of degree q with coefficients in $\bar{R}$ but not all in $\bar{M}$ and hence $t(x_1, \cdots, x_n) \notin \bar{M}^{q+1}$ so that $t(x_1, \cdots, x_n) \neq 0$ which is a contradiction since $t(x_1, \cdots, x_n) = y - y = 0$; consequently, $y \in M^q$ and $y \notin M^{q+1}$ imply that $y \notin \bar{M}^{q+1}$. Therefore, for any positive integer q, $y \in R$ and $y \notin M^q$ imply that $y \notin \bar{M}^q$; hence $y \in R$ and $y \in \bar{M}^q$ imply that $y \in M^q$. Therefore $R \cap \bar{M}^q \subset M^q$, and hence $R \cap \bar{M}^q = M^q$.

Next, let A be any ideal in R. Let $y \in (A\bar{R}) \cap R$ and let q be any positive integer; then $y \in A\bar{R}$ implies that $y = \sum a_i z_i$ with a_i in A and z_i in $\bar{R}$ so that $z_i = u_i + t_i$ with u_i in R and t_i in $\bar{M}^q$, hence $y = \sum a_i u_i + \sum a_i t_i$ with $\sum a_i u_i$ in A and $\sum a_i t_i$ in $\bar{M}^q$; since $\sum a_i t_i = y - \sum a_i u_i$ with y in R and $\sum a_i u_i$ in R we have $\sum a_i t_i \in R \cap \bar{M}^q = M^q$ so that $y = \sum a_i u_i + \sum a_i t_i \in A + M^q$. Thus $(A\bar{R}) \cap R \subset \bigcap_{q=1}^{\infty} (A + M^q) = A$, and hence $(A\bar{R}) \cap R = A$. Now it is obvious that if A is principal then so is $A\bar{R}$, so assume conversely that $A\bar{R}$ is principal and let $\bar{y}$ be a generator of $A\bar{R}$. Then it is obvious that if $\bar{y} = 0$ then $A = 0$, so assume that $\bar{y} \neq 0$ and let q be the unique integer for which $\bar{y} \in \bar{M}^q$ and $\bar{y} \notin \bar{M}^{q+1}$ so that $A\bar{R} \subset \bar{M}^q$ and $A\bar{R} \not\subset \bar{M}^{q+1}$. Since $M^{q+1}\bar{R} = \bar{M}^{q+1}$, there exists y in A such that $y \notin M^{q+1}$. Since $\bar{M}^{q+1} \cap R = M^{q+1}$, $y \notin \bar{M}^{q+1}$. Now $y = \bar{y}z$ with z in $\bar{R}$. Since $\bar{y} = \bar{M}^q$ and $y \notin \bar{M}^{q+1}$, z must be a unit in $\bar{R}$ and hence $A\bar{R} = y\bar{R}$ so that $A = (A\bar{R}) \cap R = ((yR)\bar{R}) \cap R = yR$.

Finally, take y in $K \cap \bar{R}$, then $y = z/t$ with $z \in R$ and $0 \neq t \in R$, hence $z = ty \in (t\bar{R}) \cap R = tR$, and hence $y = z/t \in R$. Therefore $K \cap \bar{R} \subset R$, and hence $K \cap \bar{R} = R$, and $K \cap \bar{M} = (K \cap \bar{M}) \cap \bar{R} = (K \cap \bar{R}) \cap \bar{M} = R \cap \bar{M} = M$.

LEMMA 3. (*A form Weierstrass Preparation Theorem*) *Let R be the formal power series ring $k[[x_1, \cdots, x_n]]$ in $x_1, \cdots, x_n$ $(n > 1)$ with coefficients in a field k, and let M be the maximal ideal in R. Let L be "the line $x_2 = \cdots = x_n = 0$." Let $S = k[[x_2, \cdots, x_n]]$, and let N be the maximal ideal in S. If y is a nonzero element in R to which L is non-tangential, then there exists a unique unit y^* in R such that $y = y'y^*$, where $y' = x_1{}^b + y_1 x_1{}^{b-1} + \cdots + y_b$, $y_i \in S$, $b = \lambda_R(y)$; furthermore $y_1, \cdots, y_b$ are necessarily in N. If z is a nonzero element in $S[x_1]$, t is a non-unit factor of z in R such that L is non-tangential to t, t^* is the unique unit in R such that $t = t't^*$, where*

$t' = x_1^d + t_1x_1^{d-1} + \cdots + t_d$, $t_i \in S$, $d = \lambda_R(t)$, *then* L *is non-tangential to* z *and* t' *divides* z *in* $S[x_1]$. *Finally, if* e *is a non-zero element of* R, *and* $u_1, \cdots, u_n$ *is a basis of* M, *and either* k *is infinite or* $\lambda_R(e) \leqq 2 = n$, *then there exist elements* a_{ij} *in* k *such that letting* $v_i = \sum a_{ij}u_j$ *we have that* $v_1, \cdots, v_n$ *is a basis of* M *and the line* $v_2 = \cdots = v_n = 0$ *is non-tangential to* e.

Proof. Follows from well known considerations; see for instance Expose X in H. Cartan's Seminar of 1951-1952. The final statement follows from the facts that the projective plane over any field has at least three rational points, and that in a projective space over an infinite field there exist rational points lying outside a given hyper-surface.

LEMMA 4. *Let* (R, M) *be a one dimensional noetherian local domain with quotient field* K. *Then the integral closure* S *of* R *in* K *is a principal ideal domain with only a finite number of prime ideals* $P_1, \cdots, P_n$ $(n > 0)$, *so that* S *is the intersection of* $R_{v_1}, \cdots, R_{v_n}$ *where* v_i *is the real discrete valuation of* K *with* $R_{v_i} = S_{P_i}$, *each* v_i *has center* M *in* R *and* $v_1, \cdots, v_n$ *are the only valuations of* K *whose valuation rings contain* R. *Furthermore, if* R *is complete then* $n = 1$.[3]

Proof. Everything except the last statement follows from [Krull 10, Section 39]. Now assume that R is complete. Let a be any element of P_1, and let $f(X) = X^m + f_1X^{m-1} + \cdots + f_m$ be the monic polynomial of least degree with coefficients f_i in R such that $f(a) = 0$. Then $a \in P_1$ and $f_i \in R \subset S$ for $i = 1, \cdots, m$ imply that $f_m \in P_1 \cap R = P$. If some f_i were not in M, then in view of the fact that S has no non-zero zero-divisors, by Hensel's Lemma applied to the monic polynomial $f(X)$ over the complete local domain R we would get a monic polynomial $g(X)$ in $R[X]$ of degree smaller than m for which $g(a) = 0$, whence $f_i \in M$ for all i. For any j, $P_j \supset P$ and hence $f_i \in P_j$ for all i; consequently $a^m \in P_j$, and hence $a \in P_j$. Therefore $P_1 \subset P_j$ for all j, and hence $n = 1$.

LEMMA 5. *Let* (R, M) *be n-dimensional regular local domain such that* R *contains a coefficient field* k, *and let* s *be a non negative integer. Then* R/M^s *can, in a natural way, be considered to be a vector space over* k, *and then its k-dimension is* $n(s) = \binom{n+s-1}{n}$, *i. e.*, (i) *there exist elements* $y_1, \cdots, y_{n(s)}$ *in* R *such that, if* $c_1, \cdots, c_{n(s)}$ *are elements in* k *with*

[3] The proof given below yields that, more generally, without the assumption of R being one dimensional, if R is complete then S has a unique maximal ideal, i. e., S is a local ring.

$c_1y_1 + \cdots + c_{n(s)}y_{n(s)} \in M^s$ then $c_1 = \cdots = c_{n(s)} = 0$; and (ii) if $z_1, \cdots, z_t$ are elements in R with $t > n(s)$, then there exist $c_1, \cdots, c_t$ in k, which are not all zero, such that $c_1z_1 + \cdots + c_tz_t$ is in M^s.

Proof. Since $k \subset R$, R as well as M^s and hence R/M^s can be considered as vector spaces over k. Fix a minimal basis $x_1, \cdots, x_n$ of M and let $y_1, \cdots, y_q$ be all distinct monomials in $x_1, \cdots, x_n$ of degree $< s$. Then $q = n(s)$ and by well known properties of regular local domains it follows that $y_1, \cdots, y_q$ form a k-basis of R modulo M^s.

LEMMA 6. *Let V and V^* be normal irreducible projective algebraic varieties over an algebraically closed ground field k, let ϕ be a regular birational map of V^* onto V, and let K' be a finite separable extension of $k(V) = k(V^*)$. If K'/V is tamely ramified then so is K'/V^*.*

Proof. Follows from Lemmas 6 and 7 of Part I and [Abhyankar 5, Proposition 1.50 of Section 7].

LEMMA 7. *Let R be a subdomain of a field K, let $x_1, \cdots, x_n$ be elements in R with $x_1 \neq 0$, let M be the ideal in R generated by $x_1, \cdots, x_n$; let $S = R[x_2/x_1, \cdots, x_n/x_1]$ and let a be any element of K. Then $a \in S$ if and only if there exists $q \geqq 0$ such that $x_1{}^q a \in M^q$, and $a \in x_1S$ if and only if there exists $q \geqq 0$ such that $x_1{}^q a \in M^{q+1}$.*

Proof. The statements being obvious for $n = 1$ we shall take $n > 1$. First assume that $a \in S$. Then $a = f(x_2/x_1, \cdots, x_n/x_1)$ with $f(Y_2, \cdots, Y_n) \in R[Y_2, \cdots, Y_n]$ so that $f(x_2/x_1, \cdots, x_n/x_1) = F(x_1, \cdots, x_n)/x_1{}^q$, where $F(X_1, \cdots, X_n)$ is a form of some degree $q \geqq 0$ with coefficients in R, and we have $x_1{}^q a = F(x_1, \cdots, x_n) \in M^q$. Conversely, assume that $x_1{}^q a \in M^q$ for some $q \geqq 0$. Then $x_1{}^q a = F(x_1, \cdots, x_n)$ where $F(X_1, \cdots, X_n)$ is a form of degree q with coefficients in R and hence $a = F(x_1, \cdots, x_n)/x_1{}^q = F(1, x_2/x_1, \cdots, x_n/x_1) \in S$. Now assume that $a \in x_1S$. Then $a = x_1b$ with $b \in S$ so that $x_1{}^q b \in M^q$ for some $q \geqq 0$, and hence $x_1{}^q a = x_1(x_1{}^q b) \in M^{q+1}$. Conversely assume that $x_1{}^q a \in M^{q+1}$ for some $q \geqq 0$. Then $x_1{}^q a = F(x_1, \cdots, x_n)$ where $F(X_1, \cdots, X_n)$ is a form of degree $q+1$ with coefficients in R and hence $a/x_1 = F(x_1, \cdots, x_n)/x_1{}^{q+1} = F(1, x_2/x_1, \cdots, x_n/x_1) \in S$ so that $a \in x_1S$.

LEMMA 8. *Let G and H be finite cyclic groups and let $\alpha \colon G \to H$ be an epimorphism. Let h be a generator of H. Then there exists a generator g of G for which $\alpha(g) = h$.*

Proof. Let m be the order of G and let n be the order of H. Then n

divides m. Let r be a generator of G. Then for some integer u, $\alpha(r^u) = h$. Let $p_1, \cdots, p_a, q_1, \cdots, q_b$ be the distinct prime divisors of m so labelled that $p_1, \cdots, p_a$ do not divide u, and $q_1, \cdots, q_b$ do divide u. Let $v = np_1 \cdots p_a$, let $w = u + v$, and let $g = r^w$. Since r is a generator of G, $\alpha(r)$ is a generator of H. Since $h = \alpha(r^u) = (\alpha(r))^u$ is also a generator of H and since H is of order n, it follows that u and n are co-prime and hence $q_1, \cdots, q_b$ do not divide n, which implies that they do not divide v, and hence they do not divide w. Also, any p_i does not divide u but does divide v, and hence does not divide w. Therefore, none of the prime factors $p_1, \cdots, p_a, q_1, \cdots, q_b$ of m divide w, and hence g is a generator of G. Since H is of order n, and n divides v, we have $\alpha(r)^v = 1$ and hence $\alpha(g) = \alpha(r)^{u+v} = \alpha(r)^u = h$.

3. **Quadratic transforms of regular local domains.** Throughout this section, (R, M) will denote a regular local domain of dimension $n > 1$, with quotient field K.

DEFINITION 1. *A local domain (R_1, M_1) with quotient field K will be called an immediate or a first quadratic transform of R if there exists $x \in M$, $x \notin M^2$ such that R_1 is the quotient ring in K of $R[Mx^{-1}]$ with respect to a maximal ideal in $R[Mx^{-1}]$ containing x,* where $R[Mx^{-1}]$ denotes the ring obtained by adjoining to R all the elements in K of the form yx^{-1} with y in M; now $x \in M$ and $x \notin M^2$ implies that there exists a minimal basis $x = x_1, x_2, \cdots, x_n$ of M and then $R[Mx^{-1}] = R[x_2/x_1, \cdots, x_n/x_1]$, and hence by Lemma 1, R_1 is a regular local domain of dimension n having center M in R. *By induction we define an m-th quadratic transform of R to be an immediate quadratic transform of an $(m-1)$-st quadratic transform of R. Also we shall call R its only 0-th quadratic transform. Finally if R^* is an m-th quadratic transform of R for some non-negative integer m, then we shall say that R^* is a quadratic transform of R.*

LEMMA 9. *Let (R^*, M^*) be an immediate quadratic transform of R, and let $z_1, \cdots, z_n$ be a minimal basis of M. Then there exists i such that $z_j/z_i \in R^*$ for $j = 1, \cdots, n$; and if we choose any one such value of i, then setting $S = R[z_1/z_i, \cdots, z_n/z_i]$ we have that $R^* \supset S$, $N = S \cap M^*$ is a maximal ideal in S containing z_i, $R^* = S_N$, $M^* = NR^*$, $MR^* = z_iR^*$, and z_i is part of a minimal basis of M^*. Also $R^* \neq R$.*

Proof. By definition there exists $x \in M$, $x \notin M^2$ such that R^* is the quotient ring of $R[Mx^{-1}]$ with respect to a maximal ideal containing x. Let v be a valuation of K having center M^* in R^*. Then $z_j/x \in R^*$ implies that

$v(z_j/x) \geqq 0$ for all j, and hence $v(x) = v(M)$. Since $z_1, \cdots, z_n$ is a basis of M, there exists i such that $v(z_i) = v(M) = v(x)$, and then $v(z_i/x) = 0$; and this, in view of the fact that z_i/x is in R^*, implies that z_i/x is a unit in R^*, and hence x/z_i is in R^*, which in turn implies that $z_j/z_i = (z_j/x)(x/z_i)$ is in R^* for all j. Now fix i such that z_j/z_i is in R^* for all j, and let $S = R[z_1/z_i, \cdots, z_n/z_i]$. Then $R^* \supset S$. Let $N = S \cap M^*$. Then canonically $R^*/M^* \supset S/N \supset R/M$; since by Lemma 1, R^*/M^* is algebraic over R/M, S/N must be algebraic over R/M. Since $z_i \in M^* \cap S = N$, by Lemma 1 we conclude that N is a maximal ideal in S. Let $R_1 = S_N$ and $M_1 = NS$. Then R^* has center M_1 in R_1. Now v also has center M_1 in R_1, and x/z_i is in R_1, and $v(x/z_i) = 0$; whence x/z_i is a unit in R_1. Hence $R[Mx^{-1}] \subset R_1$, and

$$\begin{aligned} M^* \cap R[Mx^{-1}] &= (M_v \cap R^*) \cap R[Mx^{-1}] \\ &= M_v \cap R[Mx^{-1}] = (M_v \cap R_1) \cap R[Mx^{-1}] = M_1 \cap R[Mx^{-1}]. \end{aligned}$$

Therefore $R_1 = R^*$. By Lemma 1, S/z_iS is a polynomial ring, over a field, in $n - 1 > 0$ variables, and hence contains infinitely many maximal ideals; therefore S is not a local ring. Hence $R^* \not\subset S$ and therefore $R^* \not\subset R$. Since $z_j = (z_j/z_i)z_i$ and $(z_j/z_i) \in R^*$, $MR^* = (z_1, \cdots, z_n)R^* = z_iR^*$, and by Lemma 1, z_i is part of a minimal basis of M^*.

LEMMA 10. *Let (R_1, M_1) and (R', M') be two distinct immediate quadratic transforms of R. Then there exists a minimal basis $x_1, \cdots, x_n$ of M such that $x_j/x_1 \in R_1 \cap R'$ for all j. Also, any given valuation of K can have center at the maximal ideal of at most one immediate quadratic transform of R.*

Proof. Let $y_1, \cdots, y_n$ be a minimal basis of M, let v be a valuation of K having center M_1 in R_1, and let v' be a valuation of K having center M' in R'. If for some i, $v(y_i) = v(M)$ and $v'(y_i) = v'(M)$, then, relabelling the y_j so that $y_i = y_1$ and taking $x_j = y_j$ for all j, we are through. In the contrary case, we can relabel the y_j so that $v(y_1) = v(M) \neq v(y_2)$, and $v'(y_2) = v'(M) \neq v'(y_1)$. Let $x_1 = y_1 + y_2$ and $x_j = y_j$ for all $j > 1$. Then $v(x_1) = v(M)$ and $v'(x_1) = v'(M)$, and thus the first assertion is proved. Now suppose if possible that a valuation w of K has center M_1 in R_1 and center M' in R'. Let $S = [x_2/x_1, \cdots, x_n/x_1]$. Then by Lemma 9, $S \cap M_1$ and $S \cap M'$ are distinct maximal ideals in S. However $S \cap M_1 = S \cap M_w = S \cap M'$, which is a contradiction. Hence the second assertion is proved.

LEMMA 11. *Let (R^*, M^*) be a quadratic transform of R. Then there exists a unique integer m such that R^* is an m-th quadratic transform of R;*

also there exists a unique sequence $R = R_0, R_1, \cdots, R_m = R^*$ *such that* R_i *is an immediate quadratic transform of* R_{i-1} *for* $i = 1, \cdots, m$. *Furthermore* $R_0, \cdots, R_m$ *are the only quadratic transforms of* R *which are contained in* R^*. *Also, if* v *is a valuation of* K *having center* M^* *in* R^*, *then* $R_0, \cdots, R_m$ *are the only* q*-th quadratic transforms of* R *with* $q \leqq m$, *which are contained in* R_v (*it is obvious that in each of them* v *has center at the maximal ideal*).

Proof. Follows from Lemmas 9 and 10 by induction.

DEFINITION 2. *Let* v *be a valuation of* K *having center* M *in* R *and of* R*-dimension ero. Then* by Lemma 11 above and [Abhyankar 4, Lemma 10] it follows that *for each* m *there exists a unique* m*-th quadratic transform* (R_m, M_m) *of* R *such that* v *has center* M_m *in* R_m; *and then for all* m, R_m *is an immediate quadratic transform of* R_{m-1}, *and the* R_m*-dimension of* v *is zero. We shall say that* R_m *is the* m*-th quadratic transform of* R *along* v *and that* $R = R_0, R_1, R_2, \cdots$ *is the quadratic sequence of* R *along* v.

LEMMA 12. *Let* $(\bar{R}, \bar{M})$ *be an* n*-dimensional regular local domain such that the quotient field* $\bar{K}$ *of* $\bar{R}$ *contains* K *as a subfield,* $\bar{R}$ *has center* M *in* R, $M\bar{R} = \bar{M}$, *and the natural monomorphism* $\alpha\colon R/M \to \bar{R}/\bar{M}$ *is an epimorphism. For an immediate quadratic transform* $(\bar{R}_1, \bar{M}_1)$ *of* $\bar{R}$ *set* $\tau\bar{R}_1 = \bar{R}_1 \cap K$ *and* $\tau\bar{M}_1 = \bar{M}_1 \cap K$; *then* $\bar{R}_1$ *has center* $\tau\bar{M}_1$ *in* $\tau\bar{R}_1$, $(\tau\bar{M}_1)\bar{R}_1 = \bar{M}_1$, *and the natural monomorphism* $\tau\bar{R}_1/\tau\bar{M}_1 \to \bar{R}_1/\bar{M}_1$ *is an epimorphism. Furthermore,* τ *maps the set of all immediate quadratic transforms of* $\bar{R}$ *in a one to one manner onto the set of all immediate quadratic transforms of* R.

Proof. Let $x_1, \cdots, x_n$ be a minimal basis of M, then it is also a minimal basis of $\bar{M}$. Let

$$S = R[x_2/x_1, \cdots, x_n/x_1] \text{ and } \bar{S} = \bar{R}[x_2/x_1, \cdots, x_n/x_1].$$

First observe that $(x_1\bar{S}) \cap S = x_1S$; it being obvious that $(x_1\bar{S}) \cap S \supset x_1S$ it is enough to show that $(x_1\bar{S}) \cap S \subset x_1S$, so let $y \in (x_1\bar{S}) \cap S$; then by Lemma 7 there exist non-negative integers q and t such that $x_1{}^q y \in \bar{M}^{q+1}$ and $x_1{}^t y \in M^t$, now $x_1{}^t y \in M^t$ implies that $x_1{}^t y \in R$ and hence $x_1{}^{t+q} y \in R$, also $x_1{}^q y \in \bar{M}^{q+1}$ and $x_1{}^t \in \bar{M}^t$ imply that $x_1{}^{q+t} y \in \bar{M}^{q+t+1}$ and hence by Lemma 2,

$$x_1{}^{q+t} y \in \bar{M}^{q+t+1} \cap R = M^{q+t+1}$$

and therefore by Lemma 7, $y \in x_1S$. Therefore there is a natural monomorphism $\beta\colon S/x_1S \to \bar{S}/x_1\bar{S}$ such that the diagram

$$\begin{tikzcd} \bar{S} \arrow[r, "\bar{\gamma}"] & \bar{S}/x_1\bar{S} \\ S \arrow[u, "i"] \arrow[r, "\gamma"'] & S/x_1S \arrow[u, "\beta"'] \end{tikzcd}$$

is commutative where γ and $\bar{\gamma}$ are natural epimorphisms and i is the natural injection. By Lemma 1, $(x_1S) \cap R = M$ and $(x_1\bar{S} \cap \bar{R}) = \bar{M}$, hence we have natural monomorphisms $\bar{\mu}: \bar{R}/\bar{M} \to \bar{S}/x_1\bar{S}$ and $\mu: R/M \to S/x_1S$. It is obvious that the diagram

$$\begin{tikzcd} \bar{R}/\bar{M} \arrow[r, "\bar{\mu}"] & \bar{S}/x_1\bar{S} \\ R/M \arrow[u, "\alpha"] \arrow[r, "\mu"'] & S/x_1S \arrow[u, "\beta"'] \end{tikzcd}$$

is commutative. Since α is an epimorphism and since $\bar{S}/x_1\bar{S}$ is generated over $\bar{\mu}(\bar{R}/\bar{M})$ by $\bar{\mu}(x_2/x_1), \cdots, \bar{\mu}(x_n/x_1)$, where $x_2/x_1, \cdots, x_n/x_1$ are in S, we can conclude that β is an epimorphism. Therefore, via $\bar{\gamma}^{-1}$ and $\gamma^{-1}\beta^{-1}$ we conclude that i maps the set of all maximal ideals in S containing x_1 in a one to one manner onto the set of all maximal ideals in $\bar{S}$ containing x_1; and that if $\bar{P}$ and P are corresponding members, then $P\bar{S} = \bar{P}$, and the natural monomorphism $S/P \to \bar{S}/\bar{P}$ is an epimorphism; hence $\bar{S}_{\bar{P}}$ has center PS_P in S_P, $(PS_P)\bar{S}_{\bar{P}} = \bar{P}\bar{S}_{\bar{P}}$, and the natural monomorphism $S_P/PS_P \to \bar{S}_{\bar{P}}/P\bar{S}_{\bar{P}}$ is an empimorphism. Now everything follows by applying Lemmas 9, 10, 11.

PROPOSITION 1. *Let $(\bar{R}, \bar{M})$ be an n-dimensional regular domain such that the quotient field $\bar{K}$ of $\bar{R}$ contains K as a subfield, $\bar{R}$ has center M in R, $M\bar{R} = \bar{M}$, and the natural monomorphism $R/M \to \bar{R}/\bar{M}$ is an empimorphism.*[4] *For any quadratic transform $(\bar{R}^*, \bar{M}^*)$ of $\bar{R}$ set ${}_T\bar{R}^* = \bar{R}^* \cap K$ and ${}_T\bar{M}^* = \bar{M}^* \cap K$. If $\bar{R}^*$ is an m-th quadratic transform of $\bar{R}$ then ${}_T\bar{R}^*$ is an m-th quadratic transform of R and we have: $\bar{R}^*$ has center ${}_T\bar{M}^*$ in ${}_T\bar{R}^*$; $({}_T\bar{M}^*)^q\bar{R} = \bar{M}^{*q}$ and $\bar{M}^{*q} \cap ({}_T\bar{R}^*) = ({}_T\bar{M}^*)^q$ for any non-negative integer q; $A\bar{R}^* \cap ({}_T\bar{R}^*) = A$ for any ideal A in ${}_T\bar{R}^*$, and A is principal if and only if $A\bar{R}^*$ is principal; the natural monomorphism ${}_T\bar{R}^*/{}_T\bar{M}^* \to \bar{R}^*/\bar{M}^*$ is an epimorphism; any completion of $\bar{R}^*$ is also a completion of ${}_T\bar{R}^*$; denoting by $\bar{R} = \bar{R}_0, \bar{R}_1, \cdots, \bar{R}_m = \bar{R}^*$* and $R = R_0, R_1, \cdots, R_m = {}_T\bar{R}^*$ *the unique sequences*

[4] These conditions are satisfied if $(\bar{R}, \bar{M})$ is a completion of R.

of successive immediate quadratic transforms given by Lemma 11 we have $\tau\bar{R}_i = R_i$ *for* $i = 0, 1, \cdots, m$, *and* $R_0, R_1, \cdots, R_m$ *are the only quadratic transforms of R which are contained in* $\bar{R}^*$. *Furthermore,* τ *maps the set of all transforms of* $\bar{R}$ *in a one to one manner onto the set of all quadratic transforms of R.*

Proof. Follows from Lemmas 2, 11, 12.

4. Additional considerations for quadratic transforms of two dimensional regular local domains. Throughout this section (R, M) will denote a two dimensional regular domain with quotient field K, $(\bar{R}, \bar{M})$ will denote a completion of R, and $\bar{K}$ will denote the quotient field of $\bar{R}$. Note that by [Abhyankar 4, Theorem 3], every two dimensional regular local domain with quotient field K having center M in R is a quadratic transform of R. We shall, all through the paper use this result tacitly. The use of this result can be avoided by replacing the phrase "two dimensional regular local domain with quotient field K having center M in R" by the phrase "quadratic transform of R." However the use of the first phrase will make our definitions and results sound more intrinsic.

DEFINITION 3. *Let A be a non-zero principal ideal in R, and let* (S, N) *be a two dimensional regular local domain with quotient field K having center M in R. Then* $AS = Q_1 \cap \cdots \cap Q_s \cap Q_{s+1} \cap \cdots \cap Q_t$ *where* $Q_1, \cdots, Q_t$ *are uniquely determined primary ideals whose associate prime ideals* $P_1, \cdots, P_t$ *are distinct one dimensional prime ideals in S.* Since $M \neq 0$, there are at most a finite number of one dimensional prime ideals in S which contain M. *Label the* P_j *so that* $M \not\subset P_i$ *for* $i = 1, \cdots, s$ *and* $M \subset P_i$ *for* $i = s+1, \cdots, t$. *We define*

$$S^R[A] = \textit{the } S^R\textit{-transform of } A = Q_1 \cap \cdots \cap Q_s$$

$$\mu_{S,R}(A) = \textit{the multiplicity of } A \textit{ at } S^R = \lambda_S(S^R[A]).$$

Note that $\mathrm{Rad}_S(S^R[A]) = \mathrm{Rad}_S(S^R[\mathrm{Rad}_R A])$ and $\mu_{R,R}(A) = \lambda_R(A)$, and that if $A = A_1 \cdots A_q$ where $A_1, \cdots, A_q$ are principal ideals in R then $S^R[A] = S^R[A_1] \cdots S^R[A_q]$.

LEMMA 13. *Let* (x, y) *be a basis of M. There exists a unique immediate quadratic transform* T_X *of R which does not contain* y/x. *Let H be the set of all immediate quadratic transforms of R other than* T_X. *Let* $S = R[y/x]$ *and let* τ *be the canonical homomorphism of S onto* S/xS, *let* $k = \tau(R)$ *and* $\bar{y} = \tau(y/x)$. *Then k is a field,* $\bar{y}$ *is transcendental over k and* $S/xS = k[\bar{y}]$.

Let $\bar{W}$ be the set of all maximal ideals in $[k\bar{y}]$, and let W be the set of all maximal ideals in S containing x. For w in W denote $\tau(w)$ by $\bar{w}$. Then $w \rightarrow \bar{w}$ is a one to one map of W onto $\bar{W}$, and $w \rightarrow S_w$ is a one to one map of W onto H. For each w in W fix a generator $w'(\bar{y})$ of $\bar{w}$ and let $w^*(X,Y)$ be the unique form in $k[X,Y]$ which is not divisible by X such that $w^*(1,\bar{y}) = w'(\bar{y})$; let us denote S_w by $T_{w^*(X,Y)}$. Then (i) *for any w in W, $w^*(X,Y)$ is irreducible;* (ii) *for distinct w_1 and w_2 in W, $w^*_1(X,Y)$ and $w^*_2(X,Y)$ are co-prime; and* (iii) *any irreducible form of positive degree in $k[X,Y]$ which is not divisible by X is a constant multiple of $w^*(X,Y)$ for some w in W. Let z be a non-zero element in R, let $\zeta = \lambda_R(z)$, and let $f = f(X,Y)$ be a form in $k[X,Y]$ which either equals X or equals $w^*(X,Y)$ for some w in W. Then MT_f is a one-dimensional prime ideal in T_f, any generator of MT_f is part of a minimal basis of the maximal ideal in T_f, there exists a unique principal ideal B in T_f such that $zT_f = B(MT_f)^\zeta$, and then $B \not\subset MT_f$. Furthermore $B = T_f$ if and only if the reduced R-leading form of z with respect to (x,y) is not divisible by $f(X,Y)$ in $(R/M)[X,Y]$.*

Proof. By Lemma 1, $N' = (x/y, y)S'$ is a maximal ideal containing y in $S' = R'[x/y]$, and $x/y, y$ is a minimal basis of the maximal ideal N_X in the immediate quadratic transform $T_X = S'_{N'}$ of R. Now $x/y \in N_X$ implies that $y/x \notin T_X$ and hence $S \not\subset T_X$. Next, if (T,N) is any immediate quadratic transform of R such that $y/x \notin T$, then by Lemma 9, $x/y \in T$ and, in view of the fact that $y/x \notin T$, we have that $x/y \in N$, and hence $N \cap S' \supset N'$; since N' is maximal, Lemma 9 tells us that $T = T_X$. Therefore by Lemma 9, $w \rightarrow S_w$ is a one to one map of W onto H. Via τ, we also conclude that as w runs over W, $w^*(X,Y)$ runs over mutually co-prime irreducible forms of positive degree in $k[X,Y]$ which are prime to X, and that any such form is a constant multiple of $w^*(X,Y)$ for some w in W.

It follows from Lemma 1 that $k = \tau(R)$ is a field isomorphic to R/M, and S/xS is a polynomial ring in one variable $\bar{y}$ over k. Next, $\lambda_R(z) = \zeta$ implies that $z = \phi(x,y)$ where $\phi(X,Y)$ is a form of degree ζ in X, Y with coefficients in R but not all in M, and hence denoting by $\bar{\phi}(X,Y)$ the form obtained from $\phi(X,Y)$ by reducing its coefficients modulo M, we have $\bar{\phi}(X,Y) \neq 0$. Now $z = \phi(x,y) = x^\zeta\phi(1,y/x)$, and $\phi(1,y/x) \notin xS$ because $\bar{\phi}(1,Y) \neq 0$ and because $\bar{y} = \tau(y/x)$ is transcendental over k. Fix w in W, let $f = f(X,Y) = w^*(X,Y)$ and let N_f be the maximal ideal in T_f. Since xS is a prime ideal contained in w, we have $(xT_f) \cap S = xS$, and hence $\phi(1,y/x) \notin xT_f$. By Lemma 1, $MT_f = xT_f$ and x is part of a minimal basis of N_f; hence MT_f is a one-dimensional prime ideal in T_f, and there exists a unique

principal ideal B in T_f such that $zT_f = B(MT_f)^\zeta$, namely $B = \phi(1, y/x)T_f$ and $B \not\subset MT_f$. Since $N_f \cap S = w$, $\phi(1, y/x)$ is a unit in T_f if and only if $\phi(1, y/x) \notin w$. Also, via τ, we know that $\phi(1, y/x) \notin w$ if and only if $w^*(1, \bar{y})$ does not divide $\bar{\phi}(1, \bar{y})$ in $k[\bar{y}]$, i.e., if and only if $w^*(X, Y)$ does not divide $\bar{\phi}(X, Y)$, and it is clear that $\bar{\phi}(X, Y)$ is the reduced R-leading form of z. The case of T_X is entirely similar.

LEMMA 14. *Let v be a non-real valuation of K having center M in R. Then* (i) *v is composed*[5] *with a unique real discrete valuation w of K so that $v = w \circ v^*$ where v^* is a valuation of R_w/M_w; furthermore, v^* is also real discrete and $R_w \supset R$, and denoting $M_w \cap R$ by P_w we have that P_w either equals M or is a one dimensional prime ideal in R. Now assume that $P_w \neq M$; then* (ii) *$R_w = R_{P_w}$;* (iii) *v has a unique extension $\bar{v}$ to $\bar{K}$ for which $R_{\bar{v}} \supset \bar{R}$;* (iv) *$\bar{v}$ has center $\bar{M}$ in $\bar{R}$;* (v) *$\bar{v} = \bar{w} \circ \bar{v}^*$ where $\bar{w}$ is a unique real discrete valuation of $\bar{K}$ and $\bar{v}^*$ is a unique real discrete valuation of $R_{\bar{w}}/M_{\bar{w}}$;* (vi) *$P_{\bar{w}} = \bar{R} \cap M_{\bar{w}}$ is a one-dimensional prime ideal in $\bar{R}$;* (vii) *$P_{\bar{w}} \cap R = P_w$; and* (viii) *$P_w\bar{R} = P_{\bar{w}}H$ where H is a principal ideal in $\bar{R}$. Now let P_w denote a one-dimensional prime ideal in R, and let w be the real discrete valuation of K with $R_w = R_{P_w}$. Then* (ix) *there is at least one, and at most a finite number of valuations $v^*_1, \cdots, v^*_s$ of R_w/M_w whose valuation rings contain R/P_w; furthermore,* (x) *each v^*_i is real discrete and has center M/P_w in R/P_w;* (xi) *$v_1 = w \circ v^*_1, \cdots, v_s = w \circ v^*_s$ are exactly all the distinct non-real valuations of K composed with w and having center M in R.* (xii) *If R is complete then $s = 1$; hence in the general case we may denote by $\bar{v}_i$ the unique non-real valuation of $\bar{K}$ having center $\bar{M}$ in $\bar{R}$ which is composed with $\bar{w}_i$, where $P_1, \cdots, P_t$ are the distinct one dimensional prime divisors of $P_w\bar{R}$ in $\bar{R}$, and $R_{\bar{w}_i} = \bar{R}_{P_i}$; then* (xiii) *$s = t$, and* (xiv) *we can label the $\bar{v}_j$ so that for $j = 1, \cdots, s$; $\bar{v}_j$ is the unique extension of v_j to $\bar{K}$ having center $\bar{M}$ in $\bar{R}$.*

Proof. (i, ii, ix, x, xi, xii) follow from [Abhyankar 4, Theorem 1; Abhyankar 5, Lemma 4.2 and its proof in Section 15] and Lemma 4 above. By [Abhyankar 2, Lemma 13 of Section 7] v can be extended to a valuation $\bar{v}$ of $\bar{K}$ having center $\bar{M}$ in $\bar{R}$, also any other $\bar{K}$-extension of v whose valuation ring contains $\bar{R}$ must have center $\bar{M}$ in $\bar{R}$ since $\bar{M}$ is the only ideal in $\bar{R}$ which contracts to M in R. By (i), $\bar{v} = \bar{w} \circ \bar{v}^*$ where $\bar{w}$ is a unique real discrete valuation of $\bar{K}$ and $\bar{v}^*$ is a unique real discrete valuation of $R_{\bar{w}}/M_w$. Now $R_{\bar{w}} \supset \bar{R}$ and $M_{\bar{w}} \cap \bar{R} \neq 0$; hence $M_{\bar{w}} \cap R \neq 0$; therefore $M_{\bar{w}} \cap R_v \neq 0$.

[5] For the concept, notation and results on composite valuations see [Abhyankar 5, Section 14].

Since M_v and M_w are the only non-zero prime ideals in R_v, we must have $M_{\bar{w}} \cap R_v = M_v$ or M_w. However $M_{\bar{w}} \cap R_v = M_v$ would imply that $\bar{w}$ is a $\bar{K}$-extension of v which cannot be the case since $\bar{w}$ is real while v is not; hence $M_{\bar{w}} \cap R_v = M_w$, and hence $\bar{w}$ is an extension of w to $\bar{K}$. Now $(M_{\bar{w}} \cap \bar{R}) \cap R = M_w \cap R$ is one dimensional, and hence $P_{\bar{w}} = M_{\bar{w}} \cap \bar{R}$ must be one dimensional; this proves (iv, v, vi) and now (vii, viii) follow immediately. Next, let p_i and q_i be the K-restrictions respectively of $\bar{v}_i$ and $\bar{w}_i$. Then p_i has center M in R. Also $M_{q_i} \cap R = (M_{w_i} \cap \bar{R}) \cap R = P_i \cap R = P_w$, and hence $q_i = w \neq p_i$. Since $R_w = R_{q_i} \supset R_{p_i}$, p_i is composed with w, and hence by (i, ii) we conclude that $p_i = v_j$ for some j. Thus, K-restriction gives a mapping of the set $\bar{v}_1, \cdots, \bar{v}_t$ onto the set $v_1, \cdots, v_s$; and hence the proof of (xiii, xiv) would be complete if we show that the inverse of this mapping is single valued, i.e., if we prove the uniqueness part of (iii). Assume, if possible, that v has two distinct $\bar{K}$-extensions $\bar{v}$ and $\bar{v}'$ having center $\bar{M}$ in $\bar{R}$. By [Abhyankar 4, Theorem 1], $\bar{v}$ and $\bar{v}'$ are of $\bar{R}$-dimension zero. By [Abhyankar 4, Lemma 12], there exists n such that the n-th quadratic transform $\bar{S}$ of $\bar{R}$ along $\bar{v}$ is different from the n-th quadratic transform $\bar{S}'$ of $\bar{R}$ along $\bar{v}'$. Let $S = \bar{S} \cap K$ and $S' = \bar{S}' \cap K$. Then by Proposition 1, S and S' are distinct n-th quadratic transforms of R and clearly both are along the common K-restriction v of $\bar{v}$ and $\bar{v}'$. This contradicts Lemma 11.

DEFINITION 4. Let A be a nonzero principal ideal in R and let $A = Q_1 \cap \cdots \cap Q_t$ be the decomposition of A into primary ideals such that the associated prime ideals $P_1, \cdots, P_t$ are distinct one dimensional prime ideals. Let v be a nonreal valuation of K having center M in R such that, for the real discrete valuation w of K with which v is composed, $M_w \cap R$ is a one-dimensional prime ideal in R. If $M_w \cap R \supset A$ *then we shall say that* v *is a valuation branch of* A *at* R; *note that then for a unique* i, $M_w \cap R = P_i$ *and we shall say that* Q_i *is the* v *component of* A. *By* $\mathfrak{B}(A, R)$ *we shall denote the set of all valuation branches of* A *at* R. By Lemma 14 it follows that the valuation branches of A at R are in one to one natural correspondence with the one dimensional prime divisors of $A\bar{R}$ in $\bar{R}$, i.e., "the analytic branches of A at R." Also by Lemma 14, $\mathfrak{B}(A, R)$ is a finite set and it is empty if and only if $A = R$.

LEMMA 15. *Let* (S, N) *be a two dimensional regular local domain with quotient field* K *having center* M *in* R, *and let* A *and* B *be two nonzero principal ideals in* R. *Then we have the following:* (i) $v \in \mathfrak{B}(S^R[A], S)$, *if and only if*, $v \in \mathfrak{B}(A, R)$ *and* v *has center* N *in* S; (ii) *If* A *is a product of distinct one dimensional prime ideals in* R (*i.e., if* $A \not\subset u^2R$ *for any non-*

unit u in R), then $S^R[A]$ is a product of distinct one dimensional prime ideals in S; (iii) *If A is prime one dimensional ideal in R, then $S^R[A]$ either equals S or is a prime one dimensional ideal in S;* (iv) *If $\mu_{R,R}(A) \leqq 1$, then $\mu_{S,R}(A) \leqq 1$;*[2] (v) *If A and B have no common one dimensional prime ideal factors in R, then $S^R[A]$ and $S^R[B]$ do not have any common one dimensional prime ideal factors in S. Now let (S_1, N_1) be a two dimensional regular local domain with quotient field K having center N in S; then* (vi) $S_1{}^S[S^R[A]] = S_1{}^R[A]$.

Proof. First take v in $\mathfrak{S}(S^R[A], S)$, and let w be the real discrete valuation of K with which v is composed; then $Q = M_w \cap S$ is a one dimensional prime ideal in S containing $S^R[A]$. Now $R_w \supset S \supset R$ and $P = M_w \cap R$ is a prime ideal in R. Since K is the quotient field of R, $P \neq 0$. Also, by the defienition of $S^R[A]$, $MS \not\subset Q$ and hence $P \neq M$. Therefore, P is a one dimensional prime ideal in R, and $P \supset A$ since $Q \supset S^R[A] \supset AS$; since v has center N in S, v must have center M in R; this shows that $v \in \mathfrak{S}(A, R)$. Now take v in $\mathfrak{S}(A, R)$ such that v has center N in S. Then $M_w \cap R$ is a one dimensional prime ideal in R containing A and $(M_w \cap S) \cap R = M_w \cap R$, hence $M_w \cap S$ is a one dimensional prime ideal in S containing AS but not containing MS; therefore $v \in \mathfrak{S}(S^R[A], S)$. This proves (i). If A and B are as in (v), v and v' are elements respectively of $\mathfrak{S}(S^R[A], S)$ and $\mathfrak{S}(S^R[B], S)$, and w and w' are the real discrete valuations of K with which, respectively, v and v' are composed, then as in the proof of (i), $(M_w \cap S) \cap R$ and $(M_{w'} \cap S) \cap R$ are one dimensional prime ideals in R containing A and B respectively; hence these contractions to R are distinct; hence $M_w \cap S$ and $M_{w'} \cap S$ are distinct one dimensional prime ideals in S; this shows that $\mathfrak{S}(S^R[A], S)$ and $\mathfrak{S}(S^R[B], S)$ have no elements in common; hence $S^R[A]$ and $S^R[B]$ have no common one dimensional prome ideal factor in S. This prove (v).

Now assume that A is a prime one dimensional ideal in R and $S^R[A] \neq S$. Then $\mathfrak{S}(S^R[A], S)$ contains a valuation v. Let w be the real discrete valuation of K with which v is composed. By (i), v is in $\mathfrak{S}(A, R)$ and since A is prime, $M_w \cap R = A$; hence $w(A) = 1$. Therefore $w(AS) = 1$ and hence $w(S^R[A]) \leqq 1$. However, $v \in \mathfrak{S}(S^R[A], S)$ implies that $w(S^R[A]) > 0$; hence $w(S^R[A]) = 1$, i. e., $M_w \cap S$ is the v-component of $S^R[A]$. Since A is prime, every valuation in $\mathfrak{S}(A, R)$ is composed with w; and hence by (i), so is every valuation in $\mathfrak{S}(S^R[A], S)$. Therefore $S^R[A] = M_w \cap S$, which implies $S^R[A]$ is a one dimensional prime ideal in S. This proves (iii). Now (ii) follows from (iii) and (v) in view of the fact that if $A = A \cdot \cdot \cdot A_t$ where $A_1, \cdot \cdot \cdot, A_t$ are principal ideals in R then $S^R[A] = S^R[A_1] \cdot \cdot \cdot S^R[A_t]$.

Now assume that $\mu_{R,R}(A) \leqq 1$. If $\mu_{R,R}(A) = 0$ then $A = R$ and hence $S^R[A] = S$ so that $\mu_{S,R}(A) = 0$. Now assume that $\mu_{R,R}(A) = 1$ and let x be a generator of A. Then there exists y in R such that (x, y) is a minimal basis of M. If x/y is not in S then by Lemma 9, $MS = xS$ and hence $S^R[A] = S$; if x/y is in S but not in N then by Lemma 9, $MS = yS$ and $AS = y(x/y)S = MS$ and hence $S^R[A] = S$; finally assume that x/y is in N, then $x/y \in R[x/y] \cap N$, since $(x/y, y)R[x/y]$ maps onto the prime ideal generated by the $yR[x/y]$-residue of x/y in $R[x/y]/yR[x/y]$, we conclude that $(x/y, y)$ is a prime ideal in $R[x/y]$ containing y, and hence $(x/y, y) = R[x/y] \cap N$; this implies that $(x/y, y)$ is a minimal basis of N, hence $(x/y)S$ and $MS = yS$ do not have a common one dimensional prime divisor in S, and $AS = ((x/y)y)S$; from this we conclude that $S^R[A] = (x/y)S$ and $\lambda_S(x/y) = 1$, i.e., $\mu_{S,R}(A) = 1$. This proves (iv).

LEMMA 16. *Let A be a nonzero principal ideal in R other than R. Let $v_1, \cdots, v_t$ be the distinct members of $\mathfrak{S}(A, R)$. Then $t \geqq 1$ and for each $i \leqq t$ and each $m \geqq 0$ there exists a unique m-th quadratic transform (S_{mi}, N_{mi}) of R such that v_i has center N_{mi} in S_{mi}. If S_m is any m-th quadratic transform of R, then $\mu_{S_m,R}(A) \neq 0$ if and only if $S_m = S_{mi}$ for some i.*[6]

Proof. By [Abhyankar 4, Theorem 1], each v_i is of R-dimension zero and now everything follows from Lemma 15 and Definition 2.

DEFINITION 5. Let V be a nonsingular projective algebraic surface over an algebraically closed ground field k and let P be a point on V. Then there exists a nonsingular projective algebraic surface V_1 and a regular birational map f_1 of V_1 onto V such that P is the only fundamental point of f_1^{-1}, $f_1^{-1}(P)$ is a nonsingular irreducible algebraic curve biregularly equivalent to the projective line over k, and the set of quotient rings of all the points of $f_1^{-1}(P)$ on V_1 coincides with the set of all immediate quadratic transforms of $Q(P, V)$, [See Zariski 14]; *we shall say that the pair (V_1, f_1) is an immediate quadratic transform of V with center at P. If (V_2, f_2) is an immediate quadratic transform of V_1 (for which the center may or may not be in $f_1^{-1}(P)$) then we shall say that $(V_2, f_1 f_2)$ is a second quadratic transform of V. In this manner by induction we define an m-th quadratic transform*[7] *of V and finally*

[6] We may have $S_{mi} = S_{mj}$ for $i \neq j$. More information concerning this is provided by Lemmas 13 and 14. The consequence of Lemma 16, to the effect that for each m there is at least one and at most a finite number of m-th quadratic transforms S_m of R for which $\mu_{m,R}(A) \neq 0$, can also immediately be deduced from Lemma 13.

[7] Note that if P^* is a point on an m-th quadratic transform (V^*, f) of V, then $Q(P^*, V^*)$ is an q-th quadratic transform of $Q(f(P^*), V)$ for some $q \leqq m$.

a quadratic transform is to mean an m-th quadratic transform for some m *(for* $m = 0$, $(V, 1)$ *is to be the only 0-th quadratic transform of* V*).*

Now let W *be a curve on* V *with irreducible components* $W_1, \cdots, W_t$ *and let* (V^*, f) *be a quadratic transform of* V, *let* $W^* = f^{-1}[W]$, *let* P^* *be a point of* V^* *and let* $P = f(P^*)$. *Let* $A = M(P, W, V) =$ *the ideal of* W *at* P *on* V, let B be a defining ideal of W at P on V, and let $A^* = M(P^*, W^*, V^*)$ be the ideal of W^* at P^* on V^*. Then it is clear that $f^{-1}[W_1], \cdots, f^{-1}[W_t]$ are the distinct irreducible components of W^*. *Let* $R = Q(P, V)$ *and* $S = Q(P^*, V^*)$. *We define:*

$$\lambda(W; P, V) = \textit{multiplicity of } W \textit{ at } P \textit{ on } V = \lambda_R(A)$$

$$\mu(W; P^*, V^*, f) = \textit{multiplicity of } W \textit{ at } P^* \textit{ on } V^* \textit{ for the map } f = \mu_{S,R}(A).$$

From the results of this section, we at once deduce the following: (1) $A^* = S^R[A]$, (2) $S^R[B]$ is a defining ideal of $f^{-1}[W]$ at P^* on V^*, (3) BS is a defining ideal of $f^{-1}(W)$ at P^* on V^*, (4) $\mu(W; P^*, V^*, f) = \lambda(W^*; P^*, V^*)$.

LEMMA 17. *Assume that* R *is complete, let* z *be a nonzero nonunit in* R, *and let* $\bar{\Lambda}(z) = f_1 \cdots f_t$ *be a factorization of* $\bar{\Lambda}(z)$ *into pairwise coprime forms* $f_1, \cdots, f_t$. *Then* $z = z_1 \cdots z_t$ *where* $z_1, \cdots, z_t$ *are elements of* R *with* $\bar{\Lambda}(z_i) = f_i$ *for* $i = 1, \cdots, t$. *In particular, if* $\bar{\Lambda}(z)$ *factors into two coprime linear forms, then* $z = z_1 z_2$ *where* (z_1, z_2) *is a basis of* M.

Proof. Everything will follow if we show that, $\bar{\Lambda}(z) = fg$ where f and g are coprime forms implies that z is reducible in R. Assume the contrary. Then by Lemma 14, $\Theta(zR, R)$ contains a single element v. Since f and g are coprime, by Lemma 13 there exist two distinct immediate quadratic transforms S and S_1 of R such that $S^R[zR] \neq S$ and $S_1{}^R[zR] \neq S_1$, however in view of Lemma 16 this is a contradiction. Hence the lemma is proved.

Now using Lemma 3 we shall give another proof in case R *and* R/M *have the same characteristic and either* R/M *is infinite or* $\lambda(z) \leqq 2$. Again everything will follow if assuming that $\bar{\Lambda}(z) = fg$ with coprime forms f and g we show that z is reducible. Let k be a coefficient field in R. In view of Lemma 3 we may assume that $z = x^q + f_1(y)x^{q-1} + \cdots + f_q(y)$ where x, y is a basis of M, $q = \lambda(z)$, and $f_i(y) \in S = k[[y]]$. Since $\lambda_R(z) = q$, we must have $\lambda_S(f_i(y)) \geqq i$; hence $f_i(y) = y^i g_i(y)$ with $g_i(y) \in S$. Taking R-leading forms with respect to (k, x, y) we have

$$\Lambda(z) = X^q + g_1(0)YX^{q-1} + \cdots + g_q(0)Y^q = A(X, Y)B(X, Y)$$

where A and B are nonconstant coprime forms in $k[X, Y]$ of degrees a and b

respectively, such that $a+b=q$, the coefficient of X^a in A is 1, and the coefficient of X^b in B is 1. Hence $X^q+g_1(0)X^{q-1}+\cdots+g_q(0)=A^*(X)B^*(X)$ where $A^*(X)$ and $B^*(X)$ are nonconstant monic coprime polynomials in $k[X]$ of degrees a and b respectively. Let $F(X)=X^q+g_1(Y)X^{q-1}+\cdots+g_q(Y)\in k[[Y]][X]$. Then $F(X)\equiv A^*(X)B^*(X) \pmod Y$. Therefore by Hensel's lemma applied to the polynomial $F(X)$ over the complete local domain $k[[Y]]$, we get

$$F(X)=[X^a+G_1(Y)X^{a-1}+\cdots+G_a(Y)][X^b+H_1(Y)X^{b-1}+\cdots+H_b(Y)]$$

with $a,b>0$, $a+b=q$, and $G_i(Y),H_i(Y)\in k[[Y]]$. Therefore

$$\begin{aligned} X^q+g_1(Y)YX^{q-1}+\cdots+g_q(Y)Y^q &\\ =[X^a+G_1(Y)YX^{a-1}+\cdots+G_a(Y)Y^a]&\\ \times[X^b+H_1(Y)YX^{b-1}+\cdots+H_b(Y)Y^b]& \end{aligned}$$

and hence

$$z=[x^a+G_1(y)yx^{a-1}+\cdots+G_a(y)y^a][x^b+H_1(y)x^{b-1}+\cdots+H_b(y)y^b].$$

Remark 1. Lemma 17 is false for complete regular local domains of dimension $n>2$ as can be seen from the following example: Let $x_1,\cdots,x_n$ be independent variables over a field k of characteristic p, and let $z=x_1^t+x_2^t+x_3^t+x_1x_2x_3\in S=k[[x_1,\cdots,x_n]]$ where $t>3$ and t is prime to p in case $p\neq 0$. Applying the Jacobian Criterion to the hypersurface F given by $X_1^t+X_2^t+X_3^t+X_1X_2X_3=0$ in the affine n space over the algebraic closure $\bar{k}$ of k with coordinates $X_1,\cdots,X_t$, one can verify that the singular locus of F is the $n-3$ dimensional linear space: $X_1=X_2=X_3=0$; hence F is normal everywhere, hence in particular F is normal at $X_1=\cdots=X_n=0$, and hence S/zS is an integral domain, i.e., z is irreducible in S. However $\Lambda(z)=X_1X_2X_3$ has the pairwise coprime factors X_1, X_2, X_3.

5. Normal crossings in a regular local domain.

DEFINITION 6. *Let (R,M) be a regular local domain of dimension n, let $(\bar{R},\bar{M})$ be a completion of R, and let Q be an ideal in R. We shall say that Q has an m-fold strong normal crossing at R if there exists a minimal basis $x_1,\cdots,x_n$ of M such that $Q=x_1^{u_1}\cdots x_m^{u_m}R$ with $u_i>0$;* m is then uniquely determined by Q; in fact if $y_1,\cdots,y_n$ is any other minimal basis of M such that $Q=y_1^{v_1}\cdots y_q^{v_q}R$ with $v_i>0$, then $m=q$ and after a suitable relabelling of the y_i we have $x_i=\delta_iy_i$ and $u_i=v_i$ for $i=1,\cdots,m$ where δ_i

is a unit in R.[8] *If $Q\bar{R}$ has an m-fold strong normal crossing at $\bar{R}$ then we shall say that Q has an m-fold normal crossing at R.* It is obvious that if Q has an m-fold strong normal crossing at R, then Q has an m-fold normal crossing at R. *If Q has an m-fold normal crossing (respectively: m-fold strong normal crossing) at R for some m, $(m \leqq n)$, then we shall say that Q has a normal crossing (respectively: strong normal crossing) at R.*

Now let A be a principal ideal in R. Then it is clear that A has an m-fold strong normal crossing at R if and only if $\operatorname{Rad}_R A$ has an m-fold strong normal crossing at R, and this is so if and only if there exists a minimal basis $x_1, \cdots, x_n$ of M such that $\operatorname{Rad}_R A = x_1 \cdots x_m R$. Also, A has an m-fold normal crossing at R if and only if $\operatorname{Rad}_R A$ has an m-fold normal crossing at R, which is so if and only if $\operatorname{Rad}_{\bar{R}}(A\bar{R})$ has an m-fold strong normal crossing at $\bar{R}$, and this is so if and only if there exists a minimal basis $x_1, \cdots, x_n$ of $\bar{M}$ such that $\operatorname{Rad}_{\bar{R}}(A\bar{R}) = x_1 \cdots x_m \bar{R}$. Furthermore, if R is the quotient ring of a point on an n-dimensional irreducible algebraic variety, if A is a defining ideal of a pure $(n-1)$-dimensional subvariety W of V at P, and if B is the ideal of W at P on V, then $\operatorname{Rad}_R A = B$ and $\operatorname{Rad}_{\bar{R}}(A\bar{R}) = (\operatorname{Rad}_R A)\bar{R} = B\bar{R}$. Hence in the geometric case the definitions given here coincide with those of Part I.

LEMMA 18. *Let (R, M) be a regular local domain of dimension $n > 1$, let Q be an ideal in R, and let (R^*, M^*) be a quadratic transform of R.* (i) *If Q has a normal crossing (respectively: a strong normal crossing) at R, then QR^* has a normal crossing (respectively: a strong normal crossing) at R^*.* (ii) *If Q contains an ideal having a normal crossing (respectively: a strong normal crossing) at R, then QR^* contains an ideal having a normal crossing (respectively: a strong normal crossing) at R^*.* (iii) *If Q is a principal ideal in R, then QR^* has an m-fold normal crossing (respectively: an m-fold strong normal crossing) at R if and only if* $(\operatorname{Rad}_R Q)R^*$ *has an m-fold normal crossing (respectively: an m-fold strong normal crossing) at R^*.* (iv) *If $R^* \neq R$, then MR^* has an h-fold strong normal crossing at R^* with $h > 0$.*

Proof. (iii) follows from the equality:

[8] If $q = 0$ then $Q = R$ and hence $m = 0$, now assume that $q > 0$. Then $x_1^{u_1} \cdots x_m^{u_m} \in Q \subset y_1 R$ and $y_1 R$ is prime; hence $x_i \in y_1 R$ for some $i \leqq m$. Relabel the x_j so that $x_1 \in y_1 R$. Since $x_1 \notin M^2$, we get $x_1 R = y_1 R$. Since $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ are minimal bases of M, we get $x_i \notin x_1 R$ and $y_i \notin y_1 R$ for any $i > 1$. Hence $x_1^{u_1} \cdots x_m^{u_m}$ is in $x_1^{u_1} R$ but not in $x_1^{u_1+1} R$, and $y_1^{v_1} \cdots y_q^{v_q}$ is in $y_1^{v_1} R$ but not in $y_1^{v_1+1} R$; i.e., Q is contained in $x_1^{u_1} R$ but not in $x_1^{u_1+1} R$, and Q is contained in $x_1^{v_1} R$ but not in $x_1^{v_1+1} R$. Therefore $u_1 = v_1$, hence $x_2^{u_2} \cdots x_m^{u_m} R = y_2^{v_2} \cdots y_q^{v_q}$ and we can apply induction.

$$\mathrm{Rad}_{R^*}((\mathrm{Rad}_R Q)R^*) = \mathrm{Rad}_{R^*}(QR^*);$$

(ii) follows from (i); and in view of Lemma 9, (iv) also follows from (i); also in view of Lemma 11 and Proposition 1, it is enough to prove (i) for strong normal crossings in the case when R^* is an immediate quadratic transform of R. Let then $x_1, \cdots, x_n$ be a minimal basis of M such that $Q = x_1^{u_1} \cdots x_n^{u_n} R$ with $u_i \geqq 0$. By Lemma 9, we can relabel the x_i so that $y_i = x_i/x_1 \in R^*$ for all i, and we may further relabel the y_i so that $y_2, \cdots, y_q \in M^*$ and $y_{q+1}, \cdots, y_n \notin M^*$. Let $S = R[y_2, \cdots, y_n]$, $N = S \cap M^*$, let τ be the natural homomorphism of S onto S/x_1S, let σ be the natural homomorphism of S/x_1S onto $T = S/(x_1, y_2, \cdots, y_q)S$. Since S/x_1S is a polynomial ring in the $n-1$ variables $\tau y_2, \cdots, \tau y_n$; T must be a polynomial ring in the $n-q$ variables $\sigma\tau y_{q+1}, \cdots, \sigma\tau y_n$. Also $x_1 y_2, \cdots, y_q \in M^* \cap S = N$ implies that $\sigma\tau N$ is a maximal ideal in T and hence has a basis of $n-q$ elements $\bar{z}_{q+1}, \cdots, \bar{z}_n$. Fix z_i in S such that $\sigma\tau z_i = \bar{z}_i$ for $i = q+1, \cdots, n$; let $z_1 = x_1$ and $z_i = y_i$ for $i = 2, \cdots, q$. Then $z_1, \cdots, z_n$ is a basis of N and hence a minimal basis of M^*. Since $y_{q+1}, \cdots, y_n$ are not in M^*, they are units in R^*, and we have

$$QR^* = x_1^{u_1} \cdots x_n^{u_n} R^* = z_1^{u_1 + \cdots + u_n} z_2^{u_2} \cdots z_q^{u_q} R^*.$$

LEMMA 19. *Let (R, M) be a two dimensional regular local domain with quotient field K, let A be an ideal in R, let (R^*, M^*) be a quadratic transform of R other than R, and assume that A has a normal crossing at R such that either* (i) $A\bar{R} = y_1^{u_1} y_2^{u_2} \bar{R}$ *where (y_1, y_2) is a basis of $\bar{M}$ and $u_1 \leqq 1$ and $u_2 \leqq 1$, or* (ii) *for every one dimensional prime ideal H in the immediate quadratic transform R_1 of R contained in R^* we have that the completion of R_1/H has no nonzero nilpotent elements.*[9] *Then AR^* has a strong normal crossing at R^*.*

Proof. Because of Lemma 18, we may assume that R^* is an immediate quadratic transform of R. Let $(\bar{R}^*, \bar{M}^*)$ be the completion of R^*. Let x_1, x_2 be a minimal basis of M and let y_1, y_2 be a minimal basis of $\bar{M}$ such that $A\bar{R} = y_1^{u_1} y_2^{u_2} \bar{R}$. Let k be a representative set of R/M in R. Then

[9] Note that assumption (ii) implies and is hence equivalent to the following: if a nonzero nonunit a in R_1 is not divisible by the square of any nonunit in R_1, i. e., if a is a product of pairwise coprime nonunits in R_1, then it is so also in the completion $\bar{R}_1$ of R_1. This is clear if a is irreducible in R_1 and the general case now follows from the fact that if b and c are coprime nonunits in R_1 then $(b, c)R_1$ is primary for the maximal ideal in R_1 and hence $(b, c)\bar{R}_1$ is primary for the maximal ideal in $\bar{R}_1$ and therefore b and c are coprime in $\bar{R}_1$. Also note that assumption (ii) is satisfied if R is the quotient ring of a point on an algebraic surface.

$y_i = y^*_i + a_{i1}x_1 + a_{i2}x_2$ with $y^*_i \in \bar{M}^2$ and $a_{ij} \in k$. Since (x_1, x_2) and (y_1, y_2) are both minimal bases of $\bar{M}$ we must have $\det |a_{ij}| \notin \bar{M}^2$; hence z_1, z_2 is also a minimal basis of $\bar{M}$ where $z_i = y_i - y^*_i$. Since $z_i \in R$, (z_1, z_2) is also a minimal basis of M. Relabel the y_i so that $z_2/z_1 \in R^*$. Let $w_1 = z_1$. As in the proof of Lemma 18, either $z_2/z_1 \in M^*$ in which case, setting $w_2 = z_2/z_1$, (w_1, w_2) becomes a minimal basis of M^*; or z_2/z_1 is a unit in R^* and hence also in $\bar{R}^*$, in which case there exists w_2 in R^* such that (w_1, w_2) is a minimal basis of M^*, and then $w_2 \neq z_2/z_1$. Now $y^*_i = f_i(z_1, z_2)$ where $f_i(Z_1, Z_2)$ is a form of degree 2 in Z_1, Z_2 with coefficients in $\bar{R}$, hence $y^*_i = f_i(z_1, z_2) = w_1^2 q_i$ with $q_i = f_i(1, z_2/z_1) \in \bar{R}^*$. Therefore $y_1 = z_1 + y^*_1 = w_1 + w_1^2 q_1 = w_1 \delta_1$ where $\delta_1 = 1 + w_1 q_1$ is a unit in $\bar{R}^*$; and $y_2 = z_2 + y^*_2 = w_1(z_2/z_1) + w_1^2 q_2 = w_1 w$ where $w = (z_2/z_1) + w_1 q_2$. In case z_2/z_1 is a unit in $\bar{R}^*$, w is also a unit in $\bar{R}^*$ and we have $A\bar{R}^* = w_1{}^u \bar{R}^*$ with $u = u_1 + u_2$, and hence $AR^* = w_1{}^u R^*$ because $w_1 \in R^*$. In case z_2/z_1 is in M^*, $w_2 = z_2/z_1$, and hence $w \equiv w_2 (\operatorname{mod} w_1)$, and this implies that (w_1, w) is also a basis of $\bar{M}^*$ and $A\bar{R}^* = w_1{}^u w^v \bar{R}^*$ where $v = u_2$. By Lemma 2, there exists t in AR^* such that $t = w_1{}^u w^v \delta$ where δ is a unit in $\bar{R}^*$. Since w_1 is in R^*, $w^u\delta \in \bar{R}^* \cap K$ and hence by Lemma 2, $w^v\delta \in R^*$. Now $w^v\delta R^*$ is primary and hence so is $w^v\delta\bar{R}^* = (w^v\delta\bar{R}^*) \cap R^*$; hence $w^v\delta = q^h d$ where q is an irreducible nonunit in R^*, d is a unit in R^*, and $h \geqq 0$. If $v = 0$ then $AR^* = w_1{}^u R^*$, so now assume that $v \geqq 1$. In case of assumption (i), (w_1, q) is a basis of M^* and $AR^* = w_1{}^u q R^*$, so now assume that assumption (ii) holds. Then we must have $q = w\delta'$ where δ' is a unit in $\bar{R}^*$ and hence (w_1, q) is a basis of $\bar{M}^*$ and hence also of M^* and we have $AR^* = w_1{}^u q^v R^*$.

Remark 2. Lemma 19 is false if (R, M) is a regular local domain of dimension $n > 2$, in fact it can then happen that A has a normal crossing at R and there exists an infinite sequence $R = R_0, R_1, R_2, \cdots$ of successive immediate quadratic transforms such that AR_i does not have a strong normal crossing at R_i for any i. This can be seen from the following example: Let k be a field and let $x_1, \cdots, x_n$, y be $n+1$ independent variables over k with $n > 1$. Let $x_{ij} = x_i/y^j$, $B_j = k[x_{1j}, \cdots, x_{nj}, y]$, $P_j = (x_{1j}, \cdots, x_{nj}, y)B_j$, $R_j = (B_j)_{P_j}$, $M_j = P_j R_j$. Then R_j is an immediate quadratic eransform of R_{j-1} and the completion $(\bar{R}_j, \bar{M}_j)$ of R_j is the ring of formal power series over k in the $n+1$ variables $x_{1j}, \cdots, x_{nj}, y$. Let $a = x_1{}^3 + x_2{}^3 + x_1 x_2$. Then by Lemmas 13 and 18, aR_j has a normal crossing at R_j for all j. Also $a = y^{2j}[(x_{1j}{}^3 + x_{2j}{}^3)y^j + x_{1j}x_{2j}]$; hence if we show that $(X^3 + Y^3)Z^j + XY$ is an irreducible elements of $k[X, Y, Z]$, then it will follow that $(X_{1j}{}^3 + X_{2j}{}^3)Y^j + X_{1j}X_{2j}$ is an irreducible element of $k[X_{1j}, \cdots, X_{nj}, Y]$, and hence that

$(x_{1j}{}^3 + x_{2j}{}^3)y^j + x_{1j}x_{2j}$ is an irreducible element of R_j of R_j-leading degree 2, and from this it will follow that aR_j does not have a strong normal crossing at R_j.

For $j > 0$, since $X^3 + Y^3$ and XY are coprime polynomials in X, Y, it is enough to show that $Z^j + (XY)/(X^3 + Y^3)$ is an irreducible element of $k(X, Y)[Z]$; let z be a root of this polynomial in some extension L^* of $L = k(X, Y)$, let v be the valuation of L given by the irreducible element X of $k[X, Y]$, and let v^* be an L^*-extension of v; then

$$v^*(z^j) = v^*(XY/(X^3 + Y^3)) = v^*(X),$$

i.e., $v^*(z) = v^*(X)/j$; hence $[L(z) : L] = j$ and this proves our assertion. For $j = 0$ we have the polynomial $F = X^3 + Y^3 + XY$. If it were reducible then we would have: $F = (X^2 + \alpha X + \beta)(X + \gamma)$ with α, β, γ in $k[Y]$; since $F = X^3 (\bmod Y)$, there exists a nonzero element d in k such that either (i) $\beta = dY^2$ and $\gamma = d^{-1}Y$, or (ii) $\beta = dY$ and $\gamma = d^{-1}Y^2$. Now $\alpha + \gamma$ $=$ (coefficient of X^2 in F) $= 0$, so that $\alpha = -\gamma$, and then $Y =$ (coefficient of X in F) $= \alpha\gamma + \beta = \beta - \gamma^2$; hence in case (i): $Y = dY^2 - d^{-2}Y^2$ and in case (ii): $Y = dY - d^{-2}Y^4$; both of these are contradictions; therefore F is irreducible.

Lemma 20. *Let (R, M) be a two dimensional regular local domain, let $A = Q_1 \cap \cdots \cap Q_t$ where $Q_1, \cdots, Q_t$ are primary ideals belonging to distinct one dimensional prime ideals in R. Let (S, N) be a quadratic transform of R. Assume that each Q_i has a strong normal crossing at R and that $S^R[A]$ has an m-fold normal crossing at S with $m \leqq 1$. Then AS has a strong normal crossing at S.*

Proof. Now $\mathfrak{S}(A, R)$ is the union of the pairwise disjoint sets $\mathfrak{S}(Q_1, R)$, $\cdots, \mathfrak{S}(Q_t, R)$; by Lemma 14, $\mathfrak{S}(S^R[A], S)$ contains at most one element; and by Lemma 15, $\mathfrak{S}(S^R[A], S)$ is contained in $\mathfrak{S}(A, R)$. Therefore after a suitable relabelling of the Q_i we have that $S^R[Q_i] = S$ for all $i > 1$. Hence $S^R[A] = S^R[Q_1]$. Therefore we may assume that A is primary and then the result follows from Lemma 18.

Proposition 2. *Let (R, M) be a two dimensional regular domain which is the quotient ring of a point on an algebraic or absolute*[10] *surface, and let A be a nonzero principal ideal in R. Then there exists an integer n such that for any m-th quadratic transform S of R with $m \geqq n$, AS has a strong normal crossing at S and $S^R[A]$ has an h-fold strong normal crossing at S with $h \leqq 1$.*

---

[10] For definition see [Abhyankar 4].

Proof. Let $v_1, \cdots, v_t$ be the distinct valuation branches of A at R. By [Abhyankar 4, Proposition 5], there exists an integer n_i such that for the n_i-th quadratic transform (S_i, N_i) of R along v_i there exists a basis x_i, y_i of N_i such that $(S_i)_{(x_i S_i)}$ is the valuation ring of the real discrete valuation of the quotient field of R with which v_i is composed. Invoking Lemma 12 of [Abhyankar 4], we may assume that $R_{v_i} \not\supset S_j$ whenever $i \neq j$. Then by Lemma 15, $S_i{}^R[A] = x_i{}^{u_i} S_i$, $u_i \geqq 0, 1, \cdots, t$. Hence by part (iv) of Lemma 18, each primary component of AS_i has a strong normal crossing at S_i. Let $q = \max(n_1, \cdots, n_t)$. Then by Lemmas 15 and 18, for any q-th quadratic transform T of R, each primary component of AT has a strong normal crossing at T and $T^R[A]$ is primary. Let T_i be the q-th quadratic transform of R along v_i. Replacing R by T_i and repeating the above argument, in view of Lemma 20 we conclude that there exists an integer p such that for any p-th quadratic transform T^* of any q-th quadratic transform of R we have that, AT^* has a strong normal crossing at T^* and $T^{*R}[A]$ is primary. Finally, taking $n = p + q$ and invoking Lemma 18, we are through.

6. Strength of a singularity in a two dimensional regular local domain. Throughout this section (R, M) will denote a two dimensional regular local domain with quotient field K, $(\bar{R}, \bar{M})$ will denote the completion of R, $\bar{K}$ will denote the quotient field of $\bar{R}$, and A and B will denote nonzero principal ideals in R.

DEFINITION 7. Let S be a two dimensional regular local domain with quotient field K having center M in R. We define:

$$\begin{aligned}\nu(A, B; S, R) &= \text{strength of singularity of } A \text{ on } B \text{ at } S^R \\ &= \left\{\begin{array}{l} 0 \text{ if } \mu_{S,R}(A) \leqq 1 \text{ and } BS \text{ has a} \\ \text{normal crossing at } S; \\ \tfrac{1}{2}\mu_{S,R}(A)\,(\mu_{S,R}(A) + 1) \text{ otherwise.}\end{array}\right.\end{aligned}$$

LEMMA 21. *Let $(\bar{S}, \bar{N})$ be an n-th quadratic transform of $\bar{R}$, let $S = K \cap \bar{S}$ and $N = K \cap \bar{N}$. Then* (i) *(S, N) is an n-th quadratic transform of R;* (ii) $(S^R[A])\bar{S} = \bar{S}^R[A\bar{R}]$, (iii) $\mu_{S,R}(A) = \mu_{\bar{S},\bar{R}}(A\bar{R})$, *and* (iv) $\nu(A, B; S, R) = \nu(A\bar{R}, B\bar{R}; \bar{S}, \bar{R})$.

Proof. (i) follows from Proposition 1. In the proof of (ii) the case $n = 0$ and the case $A = R$ being trivial we may assume that $n > 0$ and $A \neq R$. Then MS and $\bar{M}\bar{S}$ are principal ideals and they contain AS and $A\bar{S}$ respectively, hence we can write $AS = DE$ where D and E are principal ideals in S such that D and MS are not contained in any one dimensional prime ideal in

S and every one dimensional prime ideal in S which contains E also contains MS. Now we must have that $D+MS$ either equals S or is primary for N; in the former case we must have $\bar{S}=(D+MS)\bar{S}=D\bar{S}+\bar{M}\bar{S}$; and in the latter case, by Proposition 1, $(D+MS)\bar{S}=D\bar{S}+\bar{M}\bar{S}$ must be primary for $\bar{N}$; hence in either case, no one dimensional prime ideal in $\bar{S}$ can contain $D\bar{S}$ as well as $\bar{M}\bar{S}$. Now $A\subset M$ implies that every one dimensional prime ideal in S which contains MS also contains AS and hence it must contain E. Therefore $\mathrm{Rad}_S E=\mathrm{Rad}_S(MS)$, and hence $\mathrm{Rad}_{\bar{S}}(E\bar{S})=\mathrm{Rad}_{\bar{S}}((\mathrm{Rad}_S E)\bar{S})$ $=\mathrm{Rad}_{\bar{S}}((\mathrm{Rad}_S(MS))\bar{S})=\mathrm{Rad}_{\bar{S}}(M\bar{S})=\mathrm{Rad}_{\bar{S}}(\bar{M}\bar{S})$. Therefore every one dimensional prime ideal in $\bar{S}$ which contains $E\bar{S}$ also contains $\bar{M}\bar{S}$. Hence $(S^R[A])\bar{S}=D\bar{S}=\bar{S}^{\bar{R}}[A\bar{R}]$. This proves (ii). Now (iii) follows from (ii) in view of Proposition 1. Finally in view of Proposition 1, BS has a normal crossing at S if and only if $B\bar{S}$ has a normal crossing at $\bar{S}$ and hence (iv) follows from (iii).

DEFINITION 8. We define:

$$\nu(A,B;R)=\text{strength of singularity of } A \text{ on } B \text{ at } R=\sum_S \nu(A,B;S,R)$$

where the sum is taken over all two dimensional regular domains S with quotient field K having center M in R.

PROPOSITION 3. (*analytic invariance of the strength of a singularity*).

$$\nu(A,B;R)=\nu(A\bar{R},B\bar{R};\bar{R}).$$

Proof. Follows from Lemma 21 and Proposition 1.

PROPOSITION 4. *If R is the quotient ring of a point on an alegbraic or absolute*[10] *surface, and if A is not contained in the square of any one dimensional prime ideal in R, then $\nu(A,B;R)$ is finite.*

Proof. Follows from Proposition 2 and Lemmas 15 and 16.

LEMMA 22. *If R is the quotient ring of a point on an algebraic surface, if $\mathrm{Rad}\,A\supset\mathrm{Rad}\,B$, and if $A\neq R$; then $\nu(A,B;R)=0$, if and only if, $\lambda_R(A)=1$ and B has a strong normal crossing at R.*

Proof. The 'if' part follows from part (iv) of Lemma 15 and part (i) of Lemma 18. Now assume the $\nu(A,B;R)=0$. Then $\nu(A,B;R,R)=0$ and hence $\lambda_R(A)=1$ and B has a normal crossing at R. Let x be a generator of A. Now $B\bar{R}=x_1^u y^v\bar{R}$ where (x_1,y) is a basis of $\bar{M}$. Now $x\bar{R}$ is prime and $B\bar{R}\subset x\bar{R}$; hence either x_1 or y equals x times a unit in $\bar{R}$, say it is x_1. Then $B\bar{R}=x^u y^v\bar{R}$. By Lemma 2, there exists z in R and a unit d

in $\bar{R}$ such that $z = x^u y^v d$. Let $t_1 = z/x^u$. Then $t_1 \in \bar{R} \cap K = R$. Since $t_1\bar{R}$ is primary, so is $t_1 R = t_1 \bar{R} \cap R$, i.e., $t_1 = t^w e_1$ where t is an irreducible nonunit in R and e_1 is a unit in R. By a well known result, if P is a one dimensional prime ideal in R then $P\bar{R}$ is not contained in the square of any one dimensional prime ideal in $\bar{R}$; hence we must have $t = ye$ where e is a unit in $\bar{R}$. Therefore (x, t) is a basis of M and $BR = x^u t^v R$.

LEMMA 23. *$\nu(A, B; R) = 0$ if and only if $\nu(A, B; R, R) = 0$.*

Proof. Follows from Lemmas 15 and 18.

LEMMA 24. *There are at most a finite number of immediate quadratic transforms $S_1, \cdots, S_n$ of R such that $\mu_{S_i, R}(A) \neq 0$. If S is an immediate quadratic transform of R other than $S_1, \cdots, S_n$ then $\nu(S^R[A], BS; S) = 0$. Also*

$$\nu(A, B; R) = \nu(A, B; R, R) + \sum_{i=1}^{n} \nu(S_i^R[A], BS_i; S_i)$$

Proof. Follows from Lemmas 15 and 16.

LEMMA 25. *$\nu(A, B; R) = \nu(A, \mathrm{Rad}_R B; R)$. If S is a quadratic transform of R then $\nu(A, B; S, R) = \nu(A, \mathrm{Rad}_R B; S, R)$.*

Proof. Follows from part (iii) of Lemma 18.

DEFINITION 9. *Let V be a nonsingular projective algebraic surface over an algebraically closed ground field, let W and W^* be curves on V such that $W \subset W^*$, let P be a point on V and let $R = Q(P, V)$, $A = M(P, W, V)$, $B =$ any defining ideal of W^* at P on V. Then we define:*

$$\begin{aligned}\nu(W, W^*; P, V) &= \textit{strength of singularity of } W \\ &\quad \textit{on } W^* \textit{ at the point } P \textit{ of } V \\ &= \nu(A, B; R).\end{aligned}$$

By Lemma 25 this does not depend on which defining ideal of W^* at P on V we take for B. *Furthermore we define:*

$$\begin{aligned}\nu(W, W^*; V) &= \textit{strength of singularities of } W \textit{ on} \\ &\quad W^* \textit{ for } V \\ &= \sum_{P \in V} \nu(W, W^*; P, V).\end{aligned}$$

By Proposition 4, each term in the above summation is finite and only a finite number of terms in the summation are nonzero, for if P is not a singular point of W^*, i.e., if the multiplicity of W^* at P on V is at most 1,

then $\nu(W, W; P, V) = 0$; therefore $\nu(W, W^*; V)$ is finite. *Now let (V^*, f) be a quadratic transform of V, let $P^* \in f^{-1}(P)$, and let $R^* = Q(P^*, V^*)$. Then we define*

$$\begin{aligned}\nu(W, W^*; P^*, V^*, f) &= \textit{strength of singularity of}\\ &\quad W \textit{ on } W^* \textit{ at the point}\\ &\quad P^* \textit{ for the map } f\\ &= \nu(A, B; R^*, R).\end{aligned}$$

Again by Lemma 25 this does not depend on which defining ideal of W^* at P on V we take for B. Also from the results proved until now it follows that $\nu(W, W^*; P^*, V^*, f) = \nu(f^{-1}[W], f^{-1}(W^*); P^*, V^*)$.

LEMMA 26. *Let V be a nonsingular projective algebraic surface over an algebraically closed ground field, let W and W^* be curves on V such that $W \subset W^*$, let P be a point of V, and let (V^*, f) be an immediate quadratic transform of V with center at P. Then* (i) $\nu(W, W^*; V) = 0$ *if and only if W is nonsingular and W^* has a strong normal crossing at each point of W;* (ii) $\nu(f^{-1}[W], f^{-1}(W^*); V^*) = \nu(W, W^*; V) - \nu(W, W^*; P, V, 1)$ *where* 1 *denotes the identity map of V onto itself.*

Proof. (i) follows from Lemma 22 and (ii) follows from Lemma 24.

PROPOSITION 5. *Let V be a nonsingular projective algebraic surface over an algebraically closed ground field k, let W be an irreducible curve on V, and let W^* be a curve on V containing W. Assume that* $\dim |W| > 1 + \nu(W, W^*; V)$. *Then there exists a quadratic transform (V^*, f) of V such that $f^{-1}(W^*)$ has a strong normal crossing at each point $f^{-1}[W]$, and* $\dim |f^{-1}[W]| > 1$.

Proof. We shall apply induction on $\nu(W, W^*; V)$. If $\nu(W, W^*; V) = 0$ then in view of Lemma 26 we may take $V^* = V$; so now assume that $\nu(W, W^*; V) = n > 0$ and that the Proposition is true whenever $\nu(W, W^*; V) < n$. By Lemma 23, for some point P of V we must have $\nu(W, W; P, V, 1) \neq 0$. Let (V', g) be an immediate quadratic transform of V with center at P. Let s be the multiplicity of W at P, then $s > 0$. By Lemma 26,

$$\nu(g^{-1}[W], g^{-1}(W^*); V') = \nu(W, W^*; V) - \tfrac{1}{2}s(s+1) < n.$$

Let L and L' respectively denote the k-vector spaces of all functions u in $k(V) = k(V')$ for which, respectively, $(u) + W \geqq 0$ on V and $(u) + g^{-1}[W] \geqq 0$ on V'. Then L' is a subspace of L. Let $H = g^{-1}(P)$, $J = g^{-1}[W]$, and let v and w be the valuations of $k(V)/k$ having, respectively, H and J as

centers on V'. Let (R, M) be the quotient ring of P on V. Let $y_1, \cdots, y_t$ be arbitrary elements of L with $t > \frac{1}{2}s(s+1) = \binom{2+s-1}{2}$. Fix a generator z of $M(P, W, V)$ in R. Since $(y_i) + W \geqq 0$ on V, it follows that the polar divisor of zy_i does not go through P and hence $zy_i \in R$. By Lemma 5, there exist elements $c_1, \cdots, c_t$ in k which are not all zero such that $c_1zy_1 + \cdots + c_tzy_t \in M^s$. Now $z \in M^s$ and $z \notin M^{s-1}$, which means that $v(z) = s$ and $v(c_1zy_1 + \cdots + c_tzy_t) \geqq s$; hence $v(c_1y_1 + \cdots + c_ty_t) \geqq 0$. Since $(c_1y_1 + \cdots + c_ty_t) + W \geqq 0$ on V and, v is the only valuation of $k(V)/k$ whose center on V' is a curve and whose center on V is a point, we conclude that $(c_1y_1 + \cdots + c_ty_t) + J \geqq 0$ on V', i.e., $c_1y_1 + \cdots + c_ty_t \in L'$. This shows that $\dim_k(L'/L) \leqq \frac{1}{2}s(s+1)$ and hence $\dim L' \geqq \dim L - \frac{1}{2}s(s+1)$. Therefore

$$\begin{aligned}\dim |J| \geqq \dim |W| - \tfrac{1}{2}s(s+1) &> 1 + \nu(W, W; V) - \tfrac{1}{2}s(s+1) \\ &= 1 + \nu(J, g^{-1}(W^*); V').\end{aligned}$$

Therefore by the induction hypothesis there exists a quadratic transform (V^*, h) of V' such that $h^{-1}(g^{-1}(W^*))$ has a strong normal crossing at each point of $h^{-1}[J]$ and $\dim |h^{-1}[J]| > 1$. Let $f = gh$. Then $h^{-1}[J] = f^{-1}[W]$ and $h^{-1}(g^{-1}(W^*)) = f^{-1}(W^*)$. This completes the proof.

7. Computations of the strength of a singularity. Throughout this section, (R, M) will denote a two dimensional regular local domain with quotient field K, $(\bar{R}, \bar{M})$ will denote the completion of R, $\bar{K}$ will denote the quotient field of $\bar{R}$, A and C will denote nonzero principal ideals in R, B will denote a nonzero principal ideal in R such that $\operatorname{Rad} B \subset \operatorname{Rad} A$; since R is a unique factorization domain there then exists a unique principal ideal B' in R such that $\operatorname{Rad} B = (\operatorname{Rad} A)B'$.

Most of the considerations that follow will be independent of the choice of a particular basis of M, and also of particular generators of the various principal ideals under consideration; hence when there is no danger of confusion, we shall omit the reference to a particular basis of M, and to a generator of a principal ideal, thus for instance: by $\bar{\Lambda}(A)$ we shall mean the reduced R-leading form of some generator of A; by "$\bar{\Lambda}(A)$ and $\bar{\Lambda}(C)$ are coprime" we shall mean "$\bar{\Lambda}_{(x,y)}(A)$ and $\bar{\Lambda}_{(x,y)}(C)$ are coprime" for a basis (x, y) of M, since this is independent of the particular basis of M as long as the same basis is used in $\bar{\Lambda}(A)$ and $\bar{\Lambda}(C)$.

DEFINITION 10. *If $\Lambda(A)$ and $\Lambda(C)$ are coprime then we shall say that A and C are nontangential at R; if A and B' are nontangential at R then*

we shall say that A is nontangential on B at R. Note that A and C are nontangential at R if and only if $A\bar{R}$ and $C\bar{R}$ are nontangential at $\bar{R}$; and A is nontangential on B at R if and only if $A\bar{R}$ is nontangential on $B\bar{R}$ at $\bar{R}$. *If $\bar{\Lambda}(A)$ is a product of s pairwise coprime linear factors then we shall say that A has an s-fold ordinary point at R*; note that in this case s is the multiplicity of A at R. Again, A has an s-fold ordinary point at R if and only if $A\bar{R}$ has an ordinary s-fold point at $\bar{R}$ and by Lemma 17, this is so if and only if $A\bar{R} = P_1 \cdots P_s$ where $P_1, \cdots, P_s$ are distinct one dimensional prime ideals in $\bar{R}$ such that $\bar{\Lambda}(P_1), \cdots, \bar{\Lambda}(P_s)$ are pairwise coprime linear forms; hence if A has an s-fold ordinary point at R then A has a normal crossing at R if and only if $s \leqq 2$ (and then this normal crossing is s-fold).

Now assume that $\lambda(A) = \lambda(C) = 1$ and $A \neq C$. Then A and C are one dimensional prime ideals, $A \not\subset C = \bigcap_{n=1}^{\infty} (C + M^n)$, and $C \not\subset A = \bigcap_{n=1}^{\infty} (A + M^n)$; hence there exist unique positive integers s and t such that $A \subset (C + M^s)$, $A \not\subset (C + M^{s+1})$, $C \subset (A + M^t)$, $C \not\subset (A + M^{t+1})$. We want to show that $s = t$. Firstly, $s = 1$ if and only if $\bar{\Lambda}(A)$ and $\bar{\Lambda}(C)$ are coprime, which is so if and only if $t = 1$. Now assume that $s > 1$, then $t > 1$. Let a and c be generators of A and C respectively. Then $a = cg + f$ with $g \in R$ and $f \in M^s$. Since $\lambda(f) = s > 1$ and $\lambda(a) = \lambda(c) = 1$ we must have $\lambda(g) = 0$, i.e., g is a unit in R; hence $c = (1/g)a - (f/g)$, $1/g \in R$, and $-(f/g) \in M^s$. Hence $s \leqq t$, and similarly $t \leqq s$. Note that R/A and R/C are one dimensional regular local domains and $s = t = \lambda_{R/A}((A + C)/A) = \lambda_{R/C}((A + C)/C)$.

If $\lambda(A) = \lambda(C) = 1$, $A \neq C$, then we shall say that A and C have an $(\lambda_{R/C}((A + C)/A))$-fold contact at R; if A and B' have an s-fold contact at R then we shall say that A has an s-fold contact on B at R. Again it is clear that A and C have an s-fold contact at R if and only if $A\bar{R}$ and $C\bar{R}$ have an s-fold contact at $\bar{R}$; and A has an s-fold contact on B at R if and only if $A\bar{R}$ has an s-fold contact on $B\bar{R}$ at $\bar{R}$.

Finally if: (1) *$s = \lambda(A) > 1$,* (2) *$A\bar{R}$ is prime, and* (3) *there exists z in $\bar{R}$ with $\lambda(z) = 1$ such that $A \subset z\bar{R} + \bar{M}^{s+1}$ and $A \not\subset z\bar{R} + \bar{M}^{s+2}$, then we shall say that A has an s-fold cusp at R.* Note that quite generally for $u \in R$, $A \subset uR + M^{\lambda(A)+1}$ if and only if $\bar{\Lambda}(u)$ divides $\bar{\Lambda}(A)$. Also, if A has an s-fold cusp at R then by Lemma 17, $\bar{\Lambda}(A)$ is the s-th power of a linear form.

Now let P be a point on a projective algebraic surface V over an algebraically closed ground field k and assume that A, B, C are the ideals of curves A^*, B^*, C^* on V at P respectively. If, respectively: (1) *A and C are nontangential at R,* (2) *A is nontangential on B at R,* (3) *A has an s-fold ordinary point at R,* (4) *A and C have an s-fold contact at R,* (5) *A has an*

s-fold contact on B at R, (6) *A has an s-fold cusp at R; then we shall respectively say that:* (1) *A^* and C^* are nontangential at P on V,* $\cdots$, (6) *A^* has an s-fold cusp at P on V.* Note that in all these definitions B may be replaced by any defining ideal of B^* at P on V; and in (1), C may be replaced by any defining ideal of C^* at P on V. Note, also, that if (4) holds then from the usual definition of intersection multiplicity it follows that $s = i(A^* \cdot C^*, P; V)$. *Now assume that V is the projective plane and that A^* has an s-fold cusp at P on V, then there exists a unique line L^* on V which is called the tangent line to A^* at P such that denoting by L the ideal of L^* at P on V, we have $A \subset L + M^{s+1}$; if furthermore $A \not\subset L + M^{s+2}$ then we shall say that A^* has an ordinary s-fold cusp at P on V* (equivalently: $i(A^* \cdot L^*, P; V) = s + 1$); note that what we have called an ordinary 2-fold cusp, is in the classical literature called a simple cusp.

PROPOSITION 6. *If A and C are nontangential at R, then $\nu(A, C; R) = \nu(A, A; R) - \nu(A, A; R, R) + \nu(A, C; R, R)$. If A is nontangential on B at R, then $\nu(A, B; R) = \nu(A, A; R) - \nu(A, A; R, R) + \nu(A, B; R, R)$. If $B = B^*D$, and A and D are nontangential at R, then $\nu(A, B; R) = \nu(A, B^*; R) - \nu(A, B^*; R, R) + \nu(A, B; R, R)$. If $A = A_1 \cdots A_t$ where $A_1, \cdots, A_t$ are principal ideals in R which are pairwise nontangential at R, then $\nu(A, A; R)$ $= \sum_{i=1}^{t} \nu(A_i, A_i; R) - \sum_{i=1}^{t} \nu(A_i, A_i; R, R) + \nu(A, A; R, R)$.*

Proof. Follows from Lemmas 13 and 24.

PROPOSITION 7. (see Figure 1 in Remark 3). *If A has an ordinary s-fold point at R then $\nu(A, A; R) = \nu(A, A; R, R) = 0$ or $\frac{1}{2}s(s+1)$ according as $s = 1$ or $s > 1$.*

Proof. By Proposition 3 we may assume the R is complete. Then by Lemma 17, $A = A_1 \cdots A_t$ where $A_1, \cdots, A_t$ are principal ideals in R whose reduced R-leading forms are pairwise coprime linear forms. Let S be an immediate quadratic transform of R. If $S^R[A] = S$ then $\nu(A, A; S, R) = 0$; if $S^R[A] \neq S$ then by Lemma 13, $S^R[A] = S^R[A_i]$ for a unique i and hence $\mathrm{Rad}_S(AS) = \mathrm{Rad}_S(A_iS)$; now A_i has a normal crossing at R and hence by Lemma 18 part (i), A_iS has a normal crossing at S and hence by Lemma 18 part (iii), AS has a normal crossing at S; also $\lambda_R(A_i) = 1$ and hence by Lemma 15 part (iv), $\mu_{S,R}(A_i) \leqq 1$ and hence $\nu(A, A; S, R) = 0$. Thus $\nu(A, A; S, R) = 0$ for every immediate quadratic transform of R and hence by Lemma 24, $\nu(A, A; R) = \nu(A, A; R, R)$. If $s > 1$ then $\nu(A, A; R, R)$ $= \frac{1}{2}(s+1)$, if $s = 1$ then A has a normal crossing at R and hence $\nu(A, A; R, R) = 0$.

LEMMA 27. *Assume that A and C have an s-fold contact at R, and that R is complete and has the same characteristic as R/M. Then M has a basis (x, y) such that the line $x=0$ is nontangential to A as well as to C. Let (x, y) be any such basis of M and let k be a coefficient field in R. Then there exist unique elements a_i, c_i in k such that $a=y+a_1x+a_2x^2+\cdots$ and $c=y+c_1x+c_2x^2+\cdots$ are generators of A and C respectively. Furthermore, $a_i=c_i$ for $i=1,2,\cdots,s-1$ and $a_s\neq c_s$.*

Proof. Since $\lambda(AC)=2$, everything except the last statement follows from Lemma 3, so now we shall prove the last statement. Now $s=1$ if and only if $\Lambda(A)=y+a_1x$ and $\Lambda(c)=y+c_1x$ are coprime, which is so if and only if $a_1\neq c_1$. So now assume that $s>1$ and that $a_1=c_1$. Note that R is the formal power series ring in the two independent variables x and y over the field k. Now $a\in cR+M^s$, and hence we can write

$$y+\sum_{i>0}a_ix^i=(y+\sum_{i>0}c_ix^i)\sum_{0\leqq i+j<s}g_{i,j}x^iy^j+\sum_{i+j\geqq s}h_{i,j}x^iy^j$$

with $g_{i,j}$ and $h_{i,j}$ in k. Multiplying out the product on the right hand side, we get:

$$\begin{aligned}y+\sum_{i>0}a_ix^i=\sum_{s-1}^{s}g_{0,j-1}y^j\\ &+\sum_{i=1}^{s-1}\Big(\sum_{p=1}^{i}c_pg_{i-p,0}+\sum_{j=1}^{s-1-i}(g_{i,j-1}+\sum_{p=1}^{i}c_pg_{i-p,j})y^j\Big)x^i\\ &+\sum_{i+j\geqq s}t_{i,j}x^iy^j\end{aligned}$$

with $t_{i,j}$ in k. For $i=0,1,\cdots,s-1$ comparing coefficients of x^iy^j for $(j=0,1,\cdots,s-i-1)$ we get the following sets $I_0, I_1,\cdots,I_{s-1}$ of equations:

$$[1=g_{0,0},\text{ and for } j=1,\cdots,s-2:\ 0=g_{0,j}]\ \cdots\cdots\cdots\ I_0;$$

and

$$[a_i=\sum_{p=1}^{i}c_pg_{i-p,0},\text{ and for } j=1,\cdots,s-1-i:\ g_{i,j-1}=-\sum_{p=1}^{i}c_pg_{i-p,j}]\ \cdots I_i$$

for $i=1,2,\cdots,s-1$.

Substituting I_0 in I_1 we get: $a_1=c_1$, and for $j=1,\cdots,s-2$: $g_{1,j-1}=-c_1g_{0,j}=0$; i.e.,

$$[a_1=c_1,\text{ and for } j=0,\cdots,s-3:\ g_{1,j}=0]\ \cdots\cdots\cdots\ J_1.$$

Now substituting I_0 and J_1 in I_2 we get

$$[a_2=c_2,\text{ and for } j=0,\cdots,s-4: g_{2,j}=0]\ \cdots\cdots\cdots\ J_2.$$

Now substituting (I_0, J_1, J_2) in I_3 get J_3; and so on. This gives us: $a_i = c_i$ for $i = 1, \cdots, s-1$; $g_{0,0} = 1$; and $g_{i,j} = 0$ whenever $0 < i + j < s - 2$. This together with the fact that $a \notin cR + M^{s+1}$ now tells us that $a_s \neq c_s$.

LEMMA 28. *If A and C have an s-fold contact at R with $s > 1$, then there exists a unique immediate quadratic transform (R_1, M_1) of R for which $\mu_{R_1,R}(A) \neq 0$. Furthermore $R_1{}^R[A]$ and $R_1{}^R[C]$ have an $s-1$ fold contact at R_1.*

Proof. The first assertion follows from Lemma 13. Now let x be a generator of A. Since $\lambda_R(x) = 1$, there exists y in R such that (x, y) is a basis of M. Let $t = x/y$, $S = R[t]$ and $N = (x, t)$. Then by Lemma 13, $R_1 = S_N$, $M_1 = NR_1 = (x, t)R_1$, and $MR_1 = yR_1$. Let $A_1 = R_1{}^R[A]$ and $C_1 = R_1{}^R[C]$. Now $AR_1 = tyR_1$ and hence $A_1 = tR_1$. Let z be a generator of C. Because of symmetry in A and C, we must have $z = \tau y$, where τ is in R_1 with R_1-leading degree 1, and $C_1 = \tau R_1$. We can write $z = f + g + xh$ where $f \in M^{s+1}$, $g \in M^s$, $g \notin xR$, and h is a unit in R. Now $f = f^* y^{s+1}$ and $g = g^* y^s$, where f^* and g^* are in S; hence $\tau = f^* y^s + g^* y^{s-1} + th$; therefore $\tau \in tR_1 + M_1{}^{s-1}$. Suppose if possible that $\tau \in tR_1 + M_1{}^s$. Then $g^* y^{s-1} = t\alpha + \beta$, where $\alpha \in R_1$ and $\beta \in M_1{}^s$. Now $H = R_1/tR_1$ is a one dimensional regular local domain. Let us denote by "bars" the residue classes modulo tR_1. Then $\lambda_H(\bar{y}) = 1$, $\lambda_H(\beta) \geqq s$ and $\bar{t} = 0$. Therefore $\lambda_H(\bar{g}^*) \geqq 1$, i.e., $g^* \in tR_1$ and hence $g^* = \dfrac{tu}{v}$ where $u \in S$, $v \in S$, $v \notin N$. By Lemma 7, we can find $q > 1$ such that $uy^q \in M^q$ and $vy^q \in M^q$. Also $v \notin N$ and $y \in N$ imply that $v \notin yS$; hence by Lemma 7, $vy^q \notin M^{q+1}$. Suppose if possible that $vy^q \in xR$, then $vy^q = xw$ with $w \in M^{q-1}$, $w \notin M^q$; hence $v = (x/y)(w/y^{q-1}) = tw^*$ with $w^* \in S$; therefore $v \in tS \subset N$, which is a contradiction. Therefore $vy^q \notin xR$. Now we have: $gv = g^* y^s v = (g^* v) y^s = (tu) y^s = (ty) u y^{s-1} = xuy^{s-1}$; hence $g(vy^q) = x(uy^q) y^{s-1} \in xR$; since xR is prime and $vy^q \notin xR$ we must have $g \in xR$, which is a contradiction. This completes the proof of the lemma.

Now we shall give another proof of the second assertion of Lemma 28 using Lemma 27 in case R and R/M have the same characteristic. Without loss of generality we can assume that R is complete, and hence we may use the notation and result of Lemma 27. Since $s > 1$, $a_1 = c_1$; and replacing y by $y + a_1 x$ we may assume that $a_1 = c_1 = 0$. Let $t = y/x$. Then by Lemma 13, (x, t) is a basis of M_1. We have

$$a = y + a_2 x^2 + a_3 x^3 + \cdots = tx + a_2 x^2 + a_3 x^3 + \cdots$$
$$= x(t + a_2 x + a_3 x^2 + \cdots);$$

hence

$$R_1{}^R[A] = (t + a_2x + a_3x^2 + \cdots)R_1,$$

and similarly

$$R_1{}^R[C] = (t + c_2x + c_3x^2 + \cdots)R_1;$$

and now the assertion follows from Lemma 27.

PROPOSITION 8. (see Figure 2 in Remark 3). *If A has an s-fold contact on B, then $\nu(A,B;R) = s$ or 0 according as $s > 1$ or $s = 1$.*

Proof. If $s = 1$, then B has a normal crossing at R and $\lambda(A) = 1$, and hence $\nu(A,B;R) = 0$; so assume that $s > 1$. Now $\lambda(A) = 1$ and hence by Lemma 13 and Part (v) of Lemma 15, or also by Lemma 16, we can conclude that for each n there is a unique n-th quadratic transform R_n of R such that $A_n = R_n{}^R[A] \neq R_n$. Now R_n is an immediate quadratic transform of R_{n-1} and we have $\nu(A,B;R) = \sum_{n=0}^{\infty} \nu(A,B;R_n,R)$. By part (iv) of Lemma 15, $\nu(A,B;R_n,R) = 0$ or 1 according as BR_n does or does not have a normal crossing at R_n. By Lemma 25 we may assume that $B = AC$ where A and C have an s-fold contact at R. Let $C_n = R_n{}^R[C]$. In view of part (v) of Lemma 15, Lemma 28 tells us that A_n and C_n have an $s - i$ fold contact at R_n for $n = 0, 1, \cdots, s-1$. Hence for $n = 0, 1, \cdots, s-2$; A_nC_n and hence BR_n does not have a normal crossing at R_n. Also, A_{s-1} and C_{s-1} have a 1-fold contact at R_{s-1}; hence $A_{s-1} = xR_{s-1}$ and $C_{s-1} = yR_{s-1}$ where (x,y) is a basis of the maximal ideal in R_{s-1}. Now $s - 1 > 0$, and hence $MR_{s-1} = zR_{s-1}$ where z is a nonzero nonunit in R_{s-1}. By part (i) of Lemma 18, AR_{s-1} and CR_{s-1} have strong normal crossings at R_{s-1}, and hence we must have that $z = u^qd$, where u is an element in R_{s-1} of R_{s-1}-leading degree 1, $q > 0$, d is a unit in R_{s-1}, the reduced R_{s-1}-leading form of z is prime to the reduced R_{s-1}-leading forms of x and y, and $AR_{s-1} = xu^aR_{s-1}$ and $CR_{s-1} = yu^bR_{s-1}$ with $a, b > 0$. Let $c = a + b$. Then $BR_{s-1} = ACR_{s-1} = xyu^cR_{s-1}$. Therefore, BR_{s-1} does not have a normal crossing at R_{s-1}, and $\mathrm{Rad}_{R_{s-1}}(BR_{s-1})$ has a 3-fold ordinary point at R_{s-1}; hence as in the proof of Proposition 7, $(\mathrm{Rad}_{R_{s-1}}(BR_{s-1}))R_s$ and hence by Lemma 18, BR_s has a normal crossing at R_s. Thus BR_n has a normal crossing at R_n if and only if $n \geqq s$; hence $\nu(A,B;R) = s$.

PROPOSITION 9. *If $A\bar{R} = P_1P_2$ where P_1 and P_2 are principal ideals in $\bar{R}$ having and s-fold contact at $\bar{R}$, then $\nu(A,A;R) = 3s$.*

Proof. Without loss of generality we may assume that $\bar{R} = R$. In the

proof of Proposition 8, substitute P_1 for A and P_2 for C. Then by Lemma 13 or also by Lemma 15, for $n=0,1,\cdots,s-1$, R_n is the only n-th quadratic transform of R for which $R_n{}^R[A]\neq R_n$; and for every immediate quadratic transform S of R_{s-1}, AS has a normal crossing at S, and $S^R[A]=$ either $S^R[P_1]$ or $S^R[P_2]$; and hence by Lemma 15, $\mu_{S,R}(A)=1$. Hence by Lemmas 15 and 18, for every m-th quadratic transform T of R with $m\geqq s$, AT has a normal crossing at T and $\mu_{T,R}(A)\leqq 1$ and hence $\nu(A,A;T,R)=0$. For For $n=0,1,\cdots,s-1$; by Lemma 28, $R_n{}^R[P_1]$ and $R_n{}^R[P_2]$ have an $s-n$ fold ordinary point at R_n, and hence $\mu_{R_n,R}(A)=\mu_{R_n,R}(P_1)+\mu_{R_n,R}(P_2)$ $=1+1=2$, and therefore $\nu(A,A;R_n,R)=\frac{1}{2}(2)(2+1)=3$. Consequently $\nu(A,A;R)=\sum_{n=0}^{s-1}\nu(A,A;R_n,R)=3s$.

LEMMA 29. *If A has an s-fold cusp at R, then there is a unique immediate quadratic transform (R_1,M_1) of R for which $\mu_{R_1,R}(A)\neq 0$. Furthermore, $R_1{}^R[A]$ has an s-fold contact on AR_1 at R_1.*

Proof. By Proposition 3, we may assume that R is complete. Let z be a generator of A. Then there exists x in R with $\lambda(x)=1$ such that $z\in xR+M^{s+1}$ and $z\notin xR+M^{s+2}$. We can find y in R such that (x,y) is a basis of M. If we take leading forms with respect to (x,y) then $\bar{\lambda}(z)$ is a nonzero constant multiple of $\bar{\lambda}(x)^s$ and hence after dividing z by a suitable unit in R we can assume that $z=f+x^s$ where $f\in M^{s+1}$. We can write $f=gx+hy^{s+1}$ where $g=\sum_{i=0}^{s}g_ix^iy^{s-i}$ and $g_i,h\in R$. Since $z\notin xR+M^{s+2}$, h must be unit in R. Let $t=x/y$, $S=R[t]$, $N=(t,y)S$. Then by Lemma 13, R_1 is unique, and $R_1=S_N$, $M_1=NR_1=(t,y)R_1$, and $MR_1=yR$. Let $q=h+t\sum_{i=0}^{s}g_it^i$, $\tau=t^s+qy$, and $A_1=R_1{}^R[A]$. Since h is a unit in R, it is also a unit in R_1. Since $t\in M_1$ and $g_i\in R_1$, we conclude that q is a unit in R_1. Now

$$\begin{aligned} z=x^s+f&=x^s+x\Big(\sum_{i=0}^{s}g_ix^iy^{s-i}\Big)+hy^{s+1}\\ &=t^sy^s+ty^{s+1}\Big(\sum_{i=0}^{s}g_it^i\Big)+hy^{s+1}\\ &=t^sy^s+qy^{s+1}\\ &=y^s\tau. \end{aligned}$$

Since (t,y) is a basis of M_1, we get $t\notin yM_1$; hence $A_1=\tau R_1$ and $\mathrm{Rad}_{R_1}(AR_1)$ $=A_1(MR_1)=y\tau R_1$. Since q is a unit in R_1 and $s>1$, it follows that: MR_1 and A_1 have an n-fold contact at R_1 with $n\leqq s$. Also, the (MR_1)-residue

Figure 1.

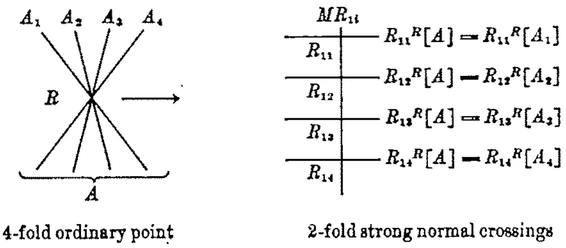


Figure 2.

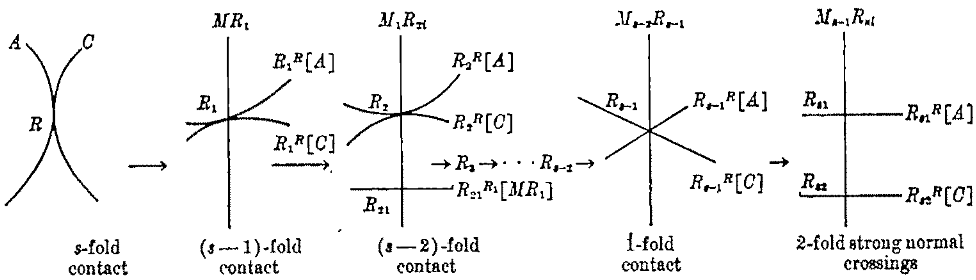


Figure 3.

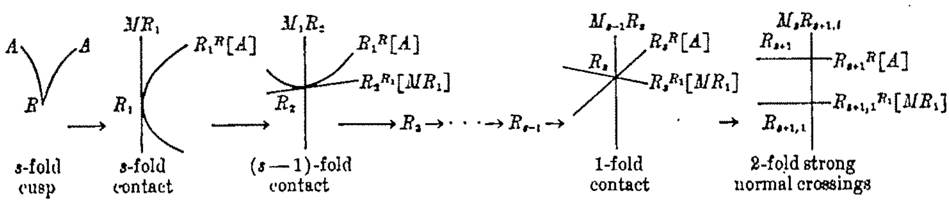
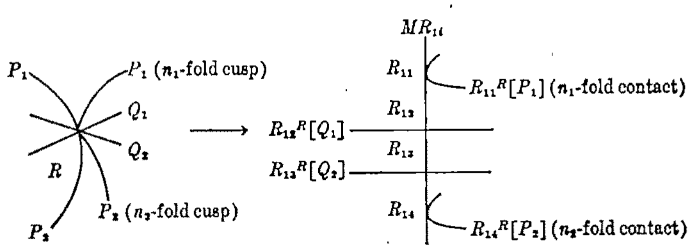


Figure 4.



class of t generates the maximal ideal in the one dimensional regular local domain R_1/MR_1, and hence the (R_1/MR_1)-leading degree of the (MR_1)-residue class of τ equals s. Therefore $n=s$.

PROPOSITION 10. (see Figure 3 in Remark 3). *If A has an s-fold cusp at R, then* $\nu(A,A;R)=\frac{1}{2}s(s+3)$.

Proof. Let R_1 be the unique immediate quadratic transform of R for which $\mu_{R_1,R}(A)\neq 0$. Then by Lemma 24:

$$\nu(A,A;R)=\nu(A,A;R,R)+\nu(R_1{}^R[A],AR_1;R_1).$$

Now $\nu(A,A;R,R)=\frac{1}{2}s(s+1)$; and by Proposition 8 and Lemma 29, $\nu(R_1{}^R[A],AR_1;R_1)=s$; hence $\mu(A,A;R)=\frac{1}{2}s(s+1)+s=\frac{1}{2}s(s+3)$.

Remark 3. It is clear that we can combine Propositions 6 to 10 and their proofs in several ways. For instance, Propositions 6 to 10 tell us that if $A=P_1\cdots P_pQ_1\cdots Q_q$ where $P_1,\cdots,P_p$, $Q_1,\cdots,Q_q$ are pairwise non-tangential principal ideals in R such that $P_1,\cdots,P_p$ have $n_1,\cdots,n_p$ fold cusps at R, respectively, and $Q_1,\cdots,Q_q$ have a simple point at R (i.e., $\lambda(Q_i)=1$), and $\lambda(A)=n_1+\cdots+n_p+q>1$, then

$$\nu(A,A;R)=\tfrac{1}{2}(n+q)(n+q+1)+n$$

where $n=n_1+\cdots+n_q$ (see Figure 4 below).

The proofs of Propositions 7, 8, 10, and the above formula, can be illustrated respectively by Figures 1 to 4 on p. 158. In these figures $(R_{h0},M_{h0})=(R_h,M_h)$, (R_{h1},M_{h1}), $(R_{h2},M_{h2}),\cdots$ denote distinct immediate quadratic transforms of (R_{h-1},M_{h-1}); also, (R_0,M_0) is the same as (R,M).

8. Local coverings at a normal crossing. Throughout this section, K will denote an n dimensional algebraic function field over an algebraically closed ground field k of characteristic p, K^* will denote a galois extension of K, V will denote a normal projective model of K/k, V^* will denote a K^*-normalization of V, ϕ will denote the rational map of V^* onto V, P will denote a simple point of V such that P is tamely ramified in K^*, P^* will denote a point in $\phi^{-1}(P)$, (R,M) will denote the quotient ring of P on V, (R^*,M^*) will denote the quotient ring of P^* on V^*, and W will denote a pure $(n-1)$ dimensional subvariety of V such that $\Delta(K^*/V)\subset W$. Now we state some propositions.

PROPOSITION 11. *If W has a normal crossing at P, then $G_i(P^*/P)$ is abelian*

PROPOSITION 12. *If W has a normal crossing at P and W_1 is an irreducible component of W having a simple point at P, then W_1 does not split locally at P^*, i.e., only one irreducible component of $\phi^{-1}(W_1)$ passes through P^*.*

PROPOSITION 13. *$\Delta(K^*/V)$ is pure $n-1$ dimensional at P.*

PROPOSITION 11α. *Proposition 11 under the assumption that W has a strong normal crossing at P.*

PROPOSITION 12α. *Proposition 12 under the assumption that W has a strong normal crossing at P.*[11]

PROPOSITION 11β. *Proposition 11 for $n=2$.*

PROPOSITION 12β. *Proposition 12 for $n=2$ (or equivalently Proposition 12α for $n=2$).*

PROPOSITION 13β. *Proposition 13 for $n=2$.*

LEMMA 30. *Proposition 12 under the assumption that $G_i(P^*/P)$ is abelian.*

Proof. Let K_i be the fixed field of $G_i(P^*/P)$, let V_i be a K_i-normalization of V, let f be the rational map of V^* onto V_i, let ϕ_i be the rational map of V_i onto V, let $P_i=f(P^*)$, and let (R_i, M_i) be the quotient ring of P_i on V_i. Let $(\bar{R}^*, \bar{M}^*)$, $(\bar{R}_i, \bar{M}_i)$ and $(\bar{R}, \bar{M})$ be the completions of R^*, R_i and R respectively. Then we can find a basis $x_1,\cdots,x_n$ of $\bar{M}$ with x_1 in R such that $x_1\cdots x_m\bar{R}=(M(W,P,V))\bar{R}$ and $x_1R=M(W_1,P,V)$. Now $x_1,\cdots,x_n$ is also a basis of $\bar{M}_i$, $x_1\in M_i$, $x_1\notin M_i^2$; hence $\phi_i^{-1}(W)$ also has a normal crossing at R_i,

$$x_1\cdots x_m\bar{R}_i=(M(\phi_i^{-1}(W),P_i,V_i))\bar{R}_i,\qquad x_1R_i=M(\phi_i^{-1}(W_1),P_i,V_i),$$

and $\phi_i^{-1}(W_1)$ has a simple point at P_i. Fix generators $q_1,\cdots,q_t$ of the ideals at P_i on V_i of the different irreducible components of $\phi^{-1}(W)$ passing through P_i. Then each q_m equals a product of a certain number of the x_j's times a unit in $\bar{R}_i$. Since $G(K^*/K_i)=G_i(R^*/R)$ is abelian and its order is prime to p in case $p\neq 0$, and since k contains all the roots of unity, we can find $z_1,\cdots,z_s$ in K^* such that $z_a^{n_a}\in K_i$ and $K^*=K_i(z_1,\cdots,z_s)$ where n_a is a positive integer which is not divisible by p in case $p\neq 0$. Multiplying z_a by a suitable element in R_i we can assume that $z_a^{n_a}\in R_i$, and since R_i is a

[11] Now it is unnecessary to say that W_1 has a simple point at P, and it is enough to say that "let W_1 be an irreducible component of W passing through P."

unique factorization domain, after dividing z_a by a suitable element in R_i we can assume that $z_a^{n_a} = d_a h_{a1}^{u_{a1}} \cdots h_{ab_a}^{u_{ab_a}}$ where $h_{a1}, \cdots, h_{ab_a}$ are pairwise coprime irreducible nonunits in R_i, d_a is a unit in R_i, and $u_{a1}, \cdots, u_{ab_a}$ are positive integers less than n_a. Then the real discrete valuation of K_i/k, whose valuation ring is the quotient ring of R_i with respect to the prime ideal $h_{ao}R_i$, is ramified in K^*; hence the irreducible $(n-1)$-dimensional subvariety of V, whose ideal at P_i on V_i is $h_{ao}R_i$, is ramified in K^* and hence it must be an irreducible component of $\phi^{-1}(W)$ passing through P_i, i.e., h_{ao} must equal some q_m times a unit in R_i, i.e., h_{ao} must equal the product of a certain number of the x_j's times a unit in $\bar{R}_i$. Let E^* and E_i be the quotient fields of $\bar{R}^*$ and $\bar{R}_i$ respectively. Then by Proposition 1 of [Abhyankar 2], $E^* = E_i(z_1, \cdots, z_s)$. Let $e = n_1 n_2 \cdots, n_s$; then e is not divisible by p in case $p \neq 0$. Let $E' = E_i(x_1^{1/e}, \cdots, x_n^{1/e})$. Since k is algebraically closed and e is not divisible by p in case $p \neq 0$, by Hensel's Lemma, every unit in E_i has all its e-th roots in E' and hence we can conclude that $E' \supset E^*$. Let $(\bar{R}', \bar{M}')$ be the integral closure of $\bar{R}^*$ in E'. Now $\bar{R}_i = k[[x_1, \cdots, x_n]]$ and hence by Lemma 5 of [Abhyankar 1], $\bar{R}' = k[[x_1^{1/e}, \cdots, x_n^{1/e}]]$, and $x_1^{1/e}, \cdots, x_n^{1/e}$ is a minimal basis of $\bar{M}'$. Hence $x_1\bar{R}'$ is a primary ideal in $\bar{R}'$, and therefore $x_1\bar{R}^*$ is a primary ideal in $\bar{R}^*$. Since $x_1\bar{R}^* \cap R^* = x_1R^*$, we conclude that x_1R^* is a primary ideal in R^*. Now x_1R^* is a defining ideal of $\phi^{-1}(W_1)$ at P^* on V^* and consequently only one irreducible component of $\phi^{-1}(W_1)$ can pass through P^*.

Remark 4. The proof of Proposition 13 given in [Abhyankar 1, Theorem 1] is incorrect, a correct proof is given in [Zariski 16] and hence now Proposition 11α follows as in [Abhyankar 1, Theorem 2] and *therefore now Proposition* 12α *follows from Proposition* 11α *and Lemma* 30. However, here we shall give a direct proof of Proposition 11β (i.e., Proposition 11 for $n = 2$) without using Proposition 13, and from Proposition 11β we shall derive Proposition 13β (i.e., Proposition 13 for $n = 2$) and this will then, by Lemma 30, give Projosition 12β. Propositions 11 and 12 were not used in Part I (See Remark 6 below) and will neither be used in this paper; we shall prove them elsewhere. Whereas Zariski's proof (of Proposition 13) uses the Jacobian theory, our proof (of Proposition 11β and hence of Proposition 13β) will, in a sense, be more geometric and will be based on the following: (1) the transform of a simple point under a quadratic transformation is a projective line, (2) a nonabelian tamely ramified covering of a projective line must have at least three branch points, (3) connectedness theorem of Zariski, and (4) the limit of a quadratic sequence of two dimen-

sional regular local domains is a valuation ring. Now for the case of an arbitrary two dimensional regular domain, (1) is still true (even for unequal characteristic), (3) has recently been proved by Chow [7], and (4) was proved by us in [4]. Hence it is expected that the proofs of Propositions 11β, 12β, 13β to be given below can probably be generalized to the case of an arbitrary two dimensional regular local domain.

LEMMA 31. *Proposition* 13β *under the assumption that* $G(K^*/K)$ *is abelian and its order is prime to* p *in case* $p \neq 0$.

Proof. If $P \notin \Delta(K^*/V)$ then there is nothing to prove, so assume that $P \in \Delta(K^*/V)$. Since K^*/K is abelian, it is a compositum of cyclic extensions and hence by Lemma 12 of Part I, there exists a field L between K and K^* such that L/K is cyclic of order n (where n is prime to p in case $p \neq 0$) such that $P \in \Delta(L/V)$. Now $L = K(z)$ with $z^n \in R$. Since R is a unique factorization domain, we can arrange matters such that $z^n = dx_1^{u_1} \cdots x_t^{u_t}$ where $x_1, \cdots, x_t$ are pairwise coprime irreducible nonunits in R, u_i are positive integers less than n, and d is a unit in R. If t were zero then the discriminant of the minimal monic polynomial of z over K would be a unit in R and hence P would unramified in L, hence $t \neq 0$. Now it is clear that the irreducible curve on V, having x_1R for its ideal at P, is ramified in L and hence it is ramified in K^*.

LEMMA 32. *Let* S *be the formal power series ring in* n *variables over* k, *let* E *be the quotient field of* S, *and let* E^* *be an abelian extension of* E *such that* $[E^* : E]$ *is prime to* p *in case* $p \neq 0$ *and such that for any minimal prime ideal* I *in* S *the real discrete valuation of* E *with valuation ring* S_I *is unramified in* E^*. *Then* $E^* = E$.

Proof. Since E^*/E is abelian, it is enough to show that if any field E' between E and E^* is cyclic over E then $E' = E$. We can find z in E' such that $E' = E(z)$ and $z^u \in S$ where u is prime to p in case $p \neq 0$. For any minimal prime ideal I in S the valuation with valuation ring S_I is unramified in E^* and hence it is unramified in E'. Since S is a unique factorization domain we can arrange matters so that z^u is not divisible by the u-th power of any irreducible nonunit in S. If z^u were then divisible by an irreducible nonunit h in S then the valuation with valuation ring S_{hS} would be ramified in E', hence z^u is a unit in S. Therefore by Hensel's Lemma $z \in E$.

LEMMA 33. *In the notation of Proposition* 11, *if* $G_i(P^*/P)$ *is metaabelian then it is abelian.*

Proof. Let K_i be the fixed field of $G_i(P^*/P)$, let $R_i = K \cap R^*$, $M_i = K \cap M^*$. Let L be a field between K^* and K_i such that K^*/K and L/K_i are abelian. Let $S = L \cap R^*$, $N = L \cap M^*$. Let $(\bar{R}^*, \bar{M}^*)$, $(\bar{S}, \bar{N})$, $(\bar{R}_i, \bar{M}_i)$ be completions, respectively, of R^*, S, R_i, and let E^*, F, E_i be the quotient fields, respectively, of $\bar{R}^*$, $\bar{S}$, $\bar{R}_i$. Now R^* is the only local ring in K^* lying above R_i and hence K_i is the splitting field of R^*/R_i as well as that of S/R_i, and L is the splitting field of R^*/S. Therefore by Lemma 7 of [A2], E^*/F, E^*/E_i, F/E_i are galois and their galois groups are isomorphic to the galois groups, respectively, of K^*/L, K^*/K_i, L/K_i. Let V_i be a K_i-normalization of V, let P_i be the point of V_i whose quotient ring on V_i is R_i, let ϕ_i be the map of V_i onto V, let $W_1, \cdots, W_t$ be the irreducible components of $\phi_i^{-1}(W)$ passing through P_i, and let $H_j = M(P_i, W_j, V_i)$. Now $\Delta(K^*/V_i) \subset \phi_i^{-1}(W)$, and as in the proof of Lemma 30 we get $(H_1 H_2 \cdots H_t)\bar{R}_i = x_1 \cdots x_m \bar{R}_i$ where $(x_1, \cdots, x_n)$ is a basis of $\bar{M}_i$, $(m \leqq n)$. Let T be the integral closure of R_i in K^* and let $\bar{T}$ be the integral closure of $\bar{R}_i$ in E^*. Let D be the ideal in R_i generated by all K_i-discriminants of all K_i bases of K^* which belong to R^* and let $\bar{D}$ be the ideal in $\bar{R}_i$ generated by all E_i-discriminants of all E_i-bases of E^* which belong to $\bar{R}^*$. Since $[E^*: E_i] = [K^*: K_i]$, we get $D \subset \bar{D}$. The only rank one prime ideals in R_i which can contain D are amongst $H_1, \cdots, H_t$ [see, Krull 9] and after a suitable relabelling of the H_j one can assume that $H_1, \cdots, H_b$ are these ideals and that $(H_1 \cdots H_b)\bar{R}_i = x_1 \cdots x_h \bar{R}_i$. Hence $D = Q_1 \cap \cdots \cap Q_b \cap J_1 \cap \cdots \cap J_a$ where Q_j is primary for H_j and J_j is primary for a prime ideal of rank > 1. Let I be a rank one prime ideal in $\bar{R}_i$ such that $\bar{D} \subset I$. Let $J = J_1 \cdots J_a$ and $Q = Q_1 \cdots Q_b$. Then $(J\bar{R}_i)(Q\bar{R}_i) = (JQ)\bar{R}_i \subset D\bar{R}_i \subset \bar{D} \subset I$. Suppose if possible that $J\bar{R}_i \subset I$. Now R_i/J is a local ring of dimension less than $n - 1$ and hence its completion $\bar{R}_i/J\bar{R}_i$ is also a local ring of dimension less than $n - 1$ [Chevalley 6, Proposition 2 of Section III]. Now $\bar{R}_i/I$ is a homomorphic image of $\bar{R}_i/J\bar{R}_i$ and hence $\bar{R}_i/I$ is a local ring of dimension less than $n - 1$ [Chevalley 6, Proposition 1 of Section III], this is a contradiction since I is a minimal ideal in $\bar{R}_i$ [Cohen 8, Theorem 8]. Therefore $J\bar{R}_i \not\subset I$ and hence $Q\bar{R}_i \subset I$. Now $Q\bar{R}_i = x_1^{u_1} \cdots x_h^{u_h} \bar{R}_i$ and hence some x_j belongs to I, i.e., $I = x_j \bar{R}_i$. Hence the minimal prime ideals in $\bar{R}_i$ which are ramified in E^* are amongst $x_1\bar{R}_i, \cdots, x_n\bar{R}_i$. Let v_i be the valuation of E_i whose valuation ring is the quotient ring of $\bar{R}_i$ with respect to $x_i\bar{R}_i$. Let e_i be the ramification index over v_i of any E^*-extension of v_i. Let $e = e_1 \cdots e_n$. Let $E^{*\prime} = E^*(x_1^{1/e}, \cdots, x_n^{1/e})$, $E'_i = E_i(x_1^{1/e}, \cdots, x_n^{1/e})$, $F' = F(x_1^{1/e}, \cdots, x_n^{1/e})$. Let $\bar{R}^*_1$, $\bar{R}'_i$, $\bar{S}'$ be the integral closures of $\bar{R}_i$ in $E^{*\prime}$, E'_i, F' respectively. Then using the techniques

of the proofs of Lemma 6 of [Abhyankar 1] and Lemma 9 of Part I, we can conclude that: (i) no minimal prime ideal in E'_i is ramified in F', (ii) no minimal prime ideal in E'_i is ramified in $E^{*\prime}$, and (iii) no minimal prime ideal in S' is ramified in $E^{*\prime}$. By [Abhyankar 1, Lemma 5],

$$E'_i = k[[x_1^{1/e}, \cdots, x_n^{1/e}]]$$

and it is clear that $E^{*\prime}/F'$ *and* F'/E'_i are abelian extensions. Hence in view of Lemma 32, (i) tells us that $F' = E'_i$ and then (iii) tells us that $E^{*\prime} = E'_i$. Therefore $E_i \subset E^* \subset E'_i$. Now E'_i/E_i is abelian and hence so is E^*/E_i. Therefore $G_i(P^*/P)$ is abelian.

LEMMA 34. *Let the assumption be as in Proposition 11β and also assume that R^* is the only local ring in K^* lying above R. Assume if possible that $G_i(P^*/P)$, i.e., $G_i(R^*/R)$, is nonabelian. Then there exists an immediate quadratic transform S of R such that for any local ring S^* in K^* lying above S, $G_i(S^*/S)$ is nonabelian.*

Proof. Note that since k is algebraically closed, $G_i(P^*/P) = G_s(P^*/P)$ and hence $G_i(P^*/P) = G(K^*/K)$. Let V_1 be an immediate quadratic transform of V with center at P, let f be the map of V_1 onto V, let V^*_1 be a K^*-normalization of V_1, let ϕ_1 be the map V^*_1 onto V_1 and let f^* be the map of V^*_1 onto V^*. Let $L = f^{-1}(P)$ and let $L^*_1, \cdots, L^*_t$ be the irreducible components of $L^* = \phi_1^{-1}(L)$. Since P^* is the only point on V^* lying above P, we have $f^{*-1}(P^*) = L^*$ and hence by the connectedness theorem [Zariski 13], L^* must be connected. Let $W_1 = f^{-1}(W)$. Then it is clear that $\Delta(V^*_1/V_1) \subset W_1$. By Lemma 19, W_1 has a strong normal crossing on V_1 at each point of L. We now have to consider two cases according as $t = 1$ or $t > 1$. *First take the case* $t > 1$. Since L^* is connected, after a suitable relabelling of the L^*_j we can assume that L^*_1 and L^*_2 have a point P^*_1 in common. Let $\phi_1(P^*_1) = P_1$. Then by Lemma 30, $G_i(P^*_1/P_1)$ is nonabelian and we can take $S = Q(P_1, V_1)$. *Next, take the case* $t = 1$. Since L does not split in K^*, the inertia field K' of L^*_1/L is a galois extension of K and L is unramified in K'. Let V'_1 be a K'-normalization of V and let ϕ' be the map of V'_1 onto V_1. Let

$$L' = \phi'^{-1}(L), \text{ and } H = Q(L, V_1)/M(L, V_1) \subset H' = Q(L', V'_1)/M(L', V'_1).$$

Then H'/H is a galois extension whose galois group is isomorphic to the galois group of K'/K. Let L_H be an H'-normalization of L and let ϕ_H be the map of L_H onto L. Then ϕ_H is the natural "lifting" to L_H of the L'-restriction of ϕ'. Hence it is clear that $\Delta(L_H/L) \subset \Delta(V'_1/V_1) \cap L$. By

Lemma 14 of Part I, $G(K^*/K')$ is cyclic and hence by Lemma 33, $G(K'/K)$ is nonabelian; hence $G(H'/H)$ is nonabelian. Now H/k is simple transcendental and L is nonsingular (i.e., L is biregularly equivalent to the projective line over k) and hence by [Abhyankar 3, Proposition 6], $\Delta(L_H/L)$ contains at least three distinct points and hence by Lemma 13, there exists a point P_1 in $\Delta(V'_1/V) \cap L$ such that $P_1 \notin f^{-1}[W]$. Therefore P_1 is an isolated point of $\Delta(V'_1/V)$ and hence by Lemma 31, for any point P^*_1 on V^*_1 lying above P_1, $G_i(P^*_1/P_1)$ must be nonabelian. Now take $S = Q(P_1, V)$.

LEMMA 35. *Same as Lemma 34 without the assumption that R^* is the only local ring in K^* lying above R.*

Proof. Let K_1 be the splitting field of R^*/R, let $R_1 = R^* \cap K_1$. Then by Lemma 34, there exists an immediate quadratic transform S_1 of R_1 such that for a local ring S^* in K^* lying above S_1, $G_i(S^*/S_1)$ is nonabelian. Let $S = S_1 \cap K$. Then by Proposition 1, S is an immediate quadratic transform of R and S_1 lies above S. Hence by Lemma 2 of Part I, $G_i(S^*/S)$ is nonabelian.

Proof of Proposition 11β. Assume, if possible, that $G_i(R^*/R)$ is nonabelian. Then by Lemma 35 there exists a sequence $R = R_0, R_1, R_2, \cdots$ of successive immediate quadratic transforms such that for each n and for any local ring R^*_n in K^* lying above R_n, $G_i(R^*_n/R_n)$ is nonabelian. By [Abhyankar 5, Lemma 4.5 of Section 15], $\bigcup_{i=0}^{\infty} R_i$ is the valuation ring of a zero dimensional valuation of K/k and hence as in the proof of [Abhyankar 5, Theorem 4.9 of Section 17] for some n, $G_i(R^*_n/R_n)$ is abelian.[12] This being a contradiction, Proposition 11β is proved.

Proof of Proposition 12β. This now follows from Proposition 11β and Lemma 30.

Proof of Proposition 13β. This now follows from Proposition 11β and Lemma 31.

Remark 5. Proposition 12β (and hence Proposition 12) is false if W_1 is only required to have a normal crossing at P instead of a strong normal crossing. We shall illustrate this by the following example. For V take the projective plane over k, let X, Y be affine coordinates in V, let (x, y) be the corresponding general point of V/k, for V^* take the surface in projective

[12] The proof of the quoted theorem applies since the order of $G_i(R^*/R)$ is prime to p in case $p \neq 0$.

three space with affine equation $Z^2 = 1 + Y$, let z be a root of $Z^2 = 1 + y$, and let ϕ be the projection along the Z-axis. Assume that $p \neq 2$. Then V^* has no singularities at finite distance and hence it is normal at finite distance. Now $K = k(x,y)$, $K^* = k(x,y,z)$, $\Delta(K^*/V) = L \cup L_\infty$ where L is the line: $1 + Y = 0$ and L_∞ is the line at infinity. Let P be the point $X = Y = 0$ and L_∞ is the line at infinity. Let P be the point $X = Y = 0$. Then above P lie the points $P_1: X = Y = 0,\ Z = 1$ and $P_2: X = Y = 0,\ Z = -1$. Let $f = y^3 + y^2 - x^2$, and let W_1 be the curve: $Y^3 + Y^2 - X^2 = 0$. Let $W = L \cup L_\infty \cup W_1$. Then $\Delta(K^*/V) \subset W$ and W has a normal crossing at P (Lemma 17) and W_1 is irreducible since $Y^3 + Y^2$ is not a square. Let (R_1, M_1) be the quotient ring of P_1 on V^*. Then R_1 is a unique factorization domain and (x,y) is a minimal basis of M_1. Now

$$f = y^3 + y^2 - x^2 = y^2(1+y) - x^2 = y^2z^2 - x^2 = (yz - x)(yz + x).$$

Since z is a unit in R_1, the reduced R_1-leading forms of $(yz - x)$ and $(yz + x)$ are coprime linear forms; hence $H_1 = (yz - x)R_1$ and $H_2 = (yz + x)R_1$ are distinct one dimensional prime ideals in R_1 and $fR_1 = H_1H_2$ and consequently $\phi^{-1}(W_1)$ splits locally at P_1 (and similarly $\phi^{-1}(W_1)$ splits locally at P_2).

Remark 6 (*Corrections to Part I*). We take this opportunity to correct some errors in Part I and to make some remarks concerning them. (1) In the third line before Proposition 1 on page 57 of Part I, the phrase "$\Delta(K^*/V)$ has a normal crossing at P" should be replaced by the phrase "$\Delta(K^*/V)$ is contained in a pure $n-1$ dimensional subvariety W of V having a strong normal crossing at P," and (2) in Proposition 2 on page 57 of Part I, the phrase "component of $\Delta(K^*/V)$" should be replaced by the phrase "component of W"; because as was shown in Remark 5 above, Proposition 2 of Part I would otherwise be false. In this modified form, Propositions 1 and 2 of Part I now follow respectively from Propositions 11α and 12α of the present paper. This modification of Proposition 2 of Part I now necessitates that the following additional changes be made in Part I: (3) Twice in the Introduction and once each in the statements of Propositions 6, 7, 8, 9 and Theorems 1, 2, 3, 4, 5 of Part I, the phrase "normal crossings" should be changed to the phrase "strong normal crossings." (4) Lines 8 and 9 in the Proof of Proposition 6 on page 74 of Part I should read "$P = \phi(P^*)$. Now $\Delta(K^*/V) \subset W$, W_j is an irreducible component of W, and W has a strong normal crossing at P." (5) In the last third and fourth lines on page 75 of Part I, "Now . . . $\Delta(K^*/V)$," should be replaced by "By (2), W has a strong normal crossing at P,". (6) After line 8 on page 90 of Part I add "and assuming that the normal crossings are strong normal crossings,".

(7) Line 30 on page 77 of Part I should read "and hence $u_{q+1}^{-1}(b_j^q)$ contains a generator of G^{q+1} and we take one such generator for b_j^{q+1}." Note that this assertion now follows from Lemma 8 of the present paper. (8) Also, referring to the fourth sentence in the proof of Lemma 32 on page 80 of Part I, the existence of v stated therein follows from Lemma 8 of the present paper in view of the fact that all the m_q-th roots of unity form a multiplicative finite cyclic group. (9) In the third line in footnote 7 on page 61 of Part I, the word "normal" is to be omitted. (10) In the first line after Lemma 4 on page 53 of Part I, the second letter K should be replaced by the letter K^*. (11) In Theorem 3 on page 79 of Part I, V should be replaced by P_n. (12) In line 7 on page 88 of Part I, the second (i) should be replaced by (ii).

(13) Proofs of Lemmas 8 and 9 of Section 2 of Part I need some clarification in the nongalois cases. Frist note that Lemmas 8 and 9 were not used in the proofs of Lemmas 10 and 11 of that section, hence the latter may be used in the proving the former. Here we shall first prove Lemma 9 and then Lemma 8.[13]

Proof of Lemma 9. First assume that K_1/V is unramified. To show that K^*_1/V^* is unramified, in view of Lemma 4 it is enough to show that any point P^* of V^* is unramified in K^*_1. Let $R^* = Q(P^*, V^*)$, let R^*_1 be a local ring in K^*_1 lying above R^*, let $R = R^*_1 \cap K$ and $R_1 = R^*_1 \cap K_1$. Let E, E_1, E^*, E^*_1 be the quotient fields of the completions of R, R_1, R^*, R^*_1 respectively. Then by Proposition 1 of [Abhyankar 2] we can conclude that E^*_1 is a compositum of E_1 and E^*, and hence $[E^*_1 : E^*] \leqq [E_1 : E]$. Since the residue fields of R and R^* are algebraically closed (they are isomorphic to k), $r(R^*_1 : R^*) = [E^*_1 : E^*]$ and $r(R_1 : R) = [E_1 : E]$. Since K_1/V is unramified, $r(R_1 : R) = 1$. Therefore $r(R^*_1 : R^*) = 1$ i.e., R^*_1/R^* is unramified. This shows that K^*_1/V^* is unramified. Now assume that K^*_1/V is tamely ramified and keep the above notation. To show that K^*_1/V^* is tamely ramified, in view of Lemmas 2 and 7 we may assume that K_1/K is galois, and in view of Lemma 11 it is enough to show that R^*_1/R^* is tamely ramified. We know that $[E_1 : E] = r(R_1/R) \not\equiv 0 \pmod p$ (for $p = 0$ there is nothing to show, so we are assuming that $p \neq 0$). Let F be a galois extension of E containing E^*_1. K_1/K is galois implies E_1/E is galois [Abhyankar 2, Section 2], i.e., $G(F/E_1)$ is a normal subgroup of $G(F/E)$, and hence $G(F/E_1) \cap G(F/E^*)$ is a normal subgroup of $G(F/E^*)$. Now $[G(F/E) : G(F/E_1)] = [E_1 : E] \not\equiv 0 \pmod p$ and by the homomorphism theorem,

[13] In these proofs, references to various lemmas are to those in Part I.

$$G(F/E^*)/(G(F/E_1) \cap G(F/E^*))$$

is isomorphic to a subgroup of $G(F/E)/G(F/E_1)$ and hence $[G(F/E^*):G(F/E_1) \cap G(F/E^*)] \not\equiv 0 \pmod p$. Now E^*_1 is the compositum of E^* and E_1 and hence $G(F/E^*_1) = G(F/E_1) \cap G(F/E^*)$. Therefore E^*_1/E^* is galois and $r[E^*_1: E^*] = [E^*_1: E^*] \not\equiv 0 \pmod p$. This completes the proof of Lemma 9. *Proof of Lemma* 8. In view of Lemmas 7 and 9, we may replace K_1 by a least galois extension L of K containing K_1 and replace K_2 by the compositum of L and K_2, and then again replace K_2 by a least galois extension of K_2 containing K_1. Thus to begin with we may assume that K_1/K and K_2/K_1 are galois. Let P be a point on V, let $R = Q(P, V)$, let R_1 be a local ring in K_1 lying above R and let R_2 be a local ring in K_2 lying above R_1. Let E, E_1, E_2 be the quotient fields of the completions of R, R_1, R_2 respectively and let E_3 be a least galois extension of E containing E_2. Then $[E_2: E_1] = r(R_2: R_1) \not\equiv 0 \pmod p$, $[E_1: E] = r(R_1: R) \not\equiv 0 \pmod p$, (for $p = 0$ there is nothing to prove, so we are assuming that $p \neq 0$), E_2/E_1 and E_1/E are galois [Abhyankar 2, Section 2]. In view of Lemma 11 it is enough to show that R_2/R is tamely ramified, i.e., $[E_3: E] \not\equiv 0 \pmod p$. Let p^q be the highest power of p that divides $[E_3: E_2]$. Then p^q is also the highest power of p that divides $[E_3: E]$ as well as the highest power of p that divides $[E_3: E_1]$. Therefore by the Sylow theorem, $G(E_3/E_2)$ contains a subgroup H of order p^q and H is necessarily a p-Sylow subgroup of $G(E_3/E)$, of $G(E_3/E_1)$, and of $G(E_3/E_2)$. Now p-Sylow subgroups of $G(E_3/E)$ are conjugates in $G(E_3/E)$ and $G(E_3/E_1)$ is a normal subgroup of $G(E_3/E)$ and hence they are all contained in $G(E_3/E_1)$ and hence they coincide with the p-Sylow subgroups of $G(E_3/E_1)$. For the same reason the p-Sylow subgroups of $G(E_3/E_1)$ are also the p-Sylow subgroups of $G(E_3/E_2)$. Consequently the subgroup T of $G(E_3/E)$ generated by all the p-Sylow subgroups of $G(E_3/E)$ belongs to $G(E_3/E_2)$. Now T is a normal subgroup of $G(E_3/E)$; however since E_3 is a least galois extension of E containing E_2, the only normal subgroup of $G(E_3/E)$ which is contained in $G(E_3/E_2)$ is the identity subgroup. Therefore $T = 1$ and hence $q = 0$. This completes the proof of Lemma 8.

9. Main results on fundamental groups. Throughout this section K will denote a two dimensional algebraic function field over an algebraically closed ground field k of characteristic p, and V will denote a nonsingular projective model of K/k, W will denote a curve on V, and $W_1, \cdots, W_t$ will denote the irreducible components of W.

PROPOSITION 14. *Let K^* be a finite separable algebraic extension of K such that K^*/V is tamely ramified and $\Delta(K^*/V) \subset W$, let V^* be a K^*-normalization of V and let ϕ be the rational map of V^* onto V. If* $\dim | W_1 | > 1 + \nu(W_1, W; V)$ *then $\phi^{-1}(W_1)$ is irreducible.*

Proof. Let K' be a least galois extension of K containing K^*, let V' be a K'-normalization of V, and let g be the rational map of V' onto V. Then by Lemmas 5 and 7 of Section 2 of Part I, K'/V is tamely ramified and $\Delta(K'/V) \subset W$. By Proposition 5 of Section 6, there exists a quadratic transform (V_1, f) of V such that $f^{-1}(W)$ has a strong normal crossing at each point of $f^{-1}[W_1]$ and $\dim | f^{-1}[W_1] | > 1$. Let V'_1 be a K'-normalization of V_1, and let g_1 be the rational map of V'_1 onto V_1. Then it is clear that $\Delta(K'/V_1) \subset f^{-1}(W)$, and by Lemma 6 of Section 2, K'/V_1 is tamely ramified. Let v be the valuation of K/k having center W_1 on V. Then v has center $f^{-1}[W_1]$ on V_1. By Proposition 5 of Section 11 of Part I, $g_1^{-1}(f^{-1}[W_1])$ is connected. Suppose if possible that $g_1^{-1}(f^{-1}(W_1])$ is reducible, then we can find two distinct irreducible components H and L of $g_1^{-1}(f^{-1}[W_1])$ which have a point P' in common. Let $P = g_1(P')$. Then $P \in f^{-1}[W_1]$ and hence $f^{-1}(W)$ has a strong normal crossing at P. Therefore by Proposition 12β of Section 8, only one irreducible component of $g_1^{-1}(f^{-1}[W_1])$ can pass through P', which is a contradiction. Therefore $g_1^{-1}(f^{-1}[W_1])$ is irreducible and hence v has only one extension to K'. Therefore v has only one extension to K^* and hence $\phi^{-1}(W_1)$ is irreducible.

PROPOSITION 15. *Assume that V is simply connected and* $\dim | W_j | > 1 + \nu(W_j, W; V)$ *for $j = 1, \cdots, t$. Let K^*/K be a galois extension such that K^*/V is tamely ramified and $\Delta(K^*/V) \subset W$. Let V^* be a K^*-normalization of V and let ϕ be the rational map of V^* onto V. Then we have the following:*

(A) *$W^*_j = \phi^{-1}(W_j)$ is irreducible for $j = 1, 2, \cdots, t$.*

(B) *The inertia group $G_i(W^*_j/W_j)$ is a cyclic normal subgroup (of order prime to p in case $p \neq 0$) of $G(K^*/K)$ for $j = 1, 2, \cdots, t$. Let a_j be generator of $G_i(W^*_j/W_j)$.*

(C) *$G(K^*/K)$ is generated by*

$$G_i(W^*_1/W_1), G_i(W^*_2/W_2), \cdots, G_i(W^*_t/W_t). \textit{ Hence}$$

(D) *$G(K^*/K)$ is generated by the t generators $a_1, a_2, \cdots, a_t$ each of which generates a normal subgroup.*

(E) *$G(K^*/K)$ is t-step nilpotent and its order is not divisible by p in case $p \neq 0$.*

(F) *If W_j and W_k have a point in common at which W has a normal crossing then a_j and a_k commute in $G(K^*/K)$.*

(G) *If W_j and W_k have a point in common at which W has a normal crossing whenever $j \neq k$, then $G(K^*/K)$ is abelian.*

Proof. Follows from the proof of Theorem 1 of Section 11 of Part I after replacing the reference there to Propositions 6 and 1 of Part I by reference to Propositions 14 and 11β of the present paper.[14]

THEOREM 1. *Assume that V is simply connected and* $\dim |W_j| > 1 + \nu(W_j, W; V)$ *for $j = 1, \cdots, t$. Then we have the following:*

(A) *$V - W$ has a tame fundamental weak parent group G generated by t generators $a_1, a_2, \cdots, a_t$ with a weak parent map f of G onto $\pi'(V - W)$ such that in each member H of $\pi'(V - W)$, a_j (i.e., the f image of a_j) generates the cyclic (and normal in H) inertia group over W_j of the unique irreducible curve corresponding to W_j on a normalization of V in the galois extension of K corresponding to H; also a_j and a_k commute in G if W_j and W_k have a point in common at which W has a normal crossing.*

(B) *Every unrestricted tame fundamental weak parent group*[15] *of $V - W$ is t-step nilpotent.*

(C) *If W_j and W_k have a point in common at which W has a normal crossing whenever $j \neq k$ then every unrestricted tame fundamental weak parent group*[15] *of $V - W$ is abelian.*

(D) $\pi^*(V - W) = \pi'(V - W)$.

Proof. Follows from the proof of[16] Theorem 2 of Section 12 of Part I after replacing the reference there to Theorem 1 of Part I by the reference to Proposition 15 above.

PROPOSITION 16. *Assume that V is simply connected and the the irreducible components W_j can be labelled so that*

---

[14] The adjective "normal crossing" in the quoted results of Part I is to be corrected to "strong normal crossing"; see Remark 6 of Section 8 of the present paper.

[15] In particular, G and the inverse limit of $\pi'(V - W)$, i.e., the galois group over K of the compositum of all the members of $\Omega_g'(V - W)$, as well as every unrestricted tame fundamental parent group of $V - W$.

[16] Note correction (7) given in Remark 6 of Section 8.

$$\dim | W_j | > \nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t)$$

for $j = 1, \cdots, t$. *Let* K^*/K be a galois extension such that K^*/V *is tamely ramified and* $\Delta(K^*/V) \subset W$. *Let* V^* *be a* K^**-normalization of* V *and let* ϕ *be the rational map of* V^* *onto* V. *Choose an irreducible component* W^*_j *of* $\phi^{-1}(W_j)$. *Let* H_j *be the subgroup of* $G(K^*/K)$ *generated by* $G_i(W^*_1/W_1)$, $\cdots, G_i(W^*_j/W_j)$, *let* K_j *be the fixed field; let* V_j *be a* K_j*-normalization of* V *and let* ϕ_j *be the rational map of* V_j *onto* V; *also set* $H_0 = 1$, $K_0 = K^*$, $V_0 = V^*$ *and* $\phi_0 = \phi$. *Then we have the following:*

(A) $\phi_j^{-1}(W_{j+1})$ *is irreducible for* $j = 0, 1, \cdots, t-1$.

(B) H_j *is a normal subgroup of* $G(K^*/K)$, *i.e.*, K_j/K *is galois for* $j = 1, \cdots, t$; *and* $H_t = G(K^*/K)$, *i.e.*, $K_t = K$. *Let* α_j *be the canonical homomorphism of* $G(K^*/K)$ *onto* $G(K_j/K)$.

(C) $\alpha_j(H_{j+1}) = G_i(\phi_j^{-1}(W_{j+1})/W_{j+1})$ *for* $j = 0, 1, \cdots, t-1$.

(D) $G_i(W^*_j/W_j)$ *and* $G_i(\phi_j^{-1}(W_j)/W_j)$ *are cyclic for* $j = 1, 2, \cdots, t$; *and if* a_j *is an element of* $G(K^*/K)$ *such that* $\alpha_j(a_j)$ *generates* $G_i(\phi_j^{-1}(W_j)/W_j)$, *in particular if* a_j *is a generator of* $G_i(W^*_j/W_j)$, *then* $a_1, \cdots, a_j$ *generate the normal subgroup* H_j *of* $G(K^*/K)$ *for* $j = 1, \cdots, t$; *and* $a_1, \cdots, a_t$ *generates* $G(K^*/K)$.

(E) $G(K^*/K)$ *is t-step solvable and its order is prime to p in case* $p \neq 0$.

Proof. (C), (D) and (F) follow at once from (A) and (B) in view of Lemmas 1 and 14 of Section 2 of Part I. We shall prove (A) and (B) by induction on t. For $t = 1$ this reduces to Proposition 15, so now assume that $t > 1$ and that (A) and (B) are true for $t-1$. By Proposition 14, $\phi_0^{-1}(W_1)$ is irreducible so that $\phi_0^{-1}(W_1) = W^*_1$ and by Lemma 1 of Section 2 of Part I, $H_1 = G_i(W^*_1/W_1)$ is a normal subgroup of $G(K^*/K)$. Let ϕ' be the rational map of V^* onto V_1, and let $W'_j = \phi'(W^*_j)$. Then W'_j is an irreducible component of $\phi_1^{-1}(W_j)$ and by Lemma 2 of Part I, $\alpha_1(H_j)$ is generated by $G_i(W'_2/W_2), \cdots, G_i(W'_j/W_j)$. By Lemmas 1 and 6 of Part I, W_1 is not ramified in K_1 and hence by Lemma 12 of Part I and Proposition 13β, $\Delta(K_1/V) \subset W_2 \cup \cdots \cup W_t$ and K_1/V is tamely ramified. Hence the induction hypothesis is satisfied if we replace K^* by K_1 and W by $W_2 \cup \cdots \cup W_t$; hence we have that $\phi_j^{-1}(W_{j+1})$ is irreducible and $\alpha_1(H_j)$ is a normal subgroup of $G(K_1/K)$ for $j = 1, \cdots, t-1$, and that $\alpha_1(H_t) = G(K_1/K)$. From this (A) and (B) follow at once. Thus the induction is complete and the proposition is proved.

THEOREM 2. *Assume that V is simply connected and that the irreducible components W_j can be labelled so that* $\dim |W_j| > 1 + \nu(W_j, W_j \cup \cdots \cup W_t)$ *for $j = 1, \cdots, t$. Then we have the following:*

(A) *$V - W$ has a tame fundamental weak parent group G generated by t generators $a_1, \cdots, a_t$ with a weak parent map f of G onto $\pi'(V - W)$ such that for each member H of $\pi'(V - W)$ after denoting the corresponding member of $\Omega'_g(V - W)$ by K^*, a K^*-normalization of V by V^* and the rational map of V^* onto V by ϕ we have the following:* (1) *for some irreducible component W^*_j of $\phi^{-1}(W_j)$, $f_H(a_j)$ is a generator of $G_i(W^*_j/W_j)$ for $j = 1, \cdots, t$;* (2) *$a_1, \cdots, a_j$ generate a normal subgroup H_j of $H = G(K^*/K)$ for $j = 1, \cdots, t$ and $H_t = H$;* (3) *if we denote the fixed field of H_j by K_j, a K_j-normalization of V by V_j, the rational map of V_j onto V by ϕ_j, the natural homomorphism of H onto $G(K_j/K)$ by α_j and H by H_0, then $\phi_j^{-1}(W_{j+1})$ is irreducible and $\alpha_j(H_{j+1}) = G_i(\phi_j^{-1}(W_{j+1})/W_{j+1})$ for $j = 0, 1, \cdots, t-1$.*

(B) *Every unrestricted tame fundamental weak parent group*[15] *of $V - W$ is t-step solvable.*

(C) $\pi^*(V - W) = \pi'(V - W)$.

Proof. Replacing the reference to[14] Theorem 1 of Part I by reference to the above Proposition 16, from any one of the three proofs (i), (ii), (iii) given in Theorem 2 of Section 12 of Part I of the initial italicized assertion in the proof of that theorem we conclude that $\pi'(V - W)$ contains an ascending cofinal sequence, $1 = G^1 < G^2 < \cdots$. Let $K = K^1 < K^2 < \cdots$ be the corresponding galois extension of K with $G(K^q/K) = G^q$. Let $v_j = v_j^1$ be the real discrete valuation of K/k having center W_j on V. Fix an arbitrary K^2-extension v_j^2 of v_j^1, then fix an arbitrary K^3-extension v_j^3 of v_j^2, and so on, thus getting a K^q-extension v_j^q of v_j for all q such that v_j^b is a K^b-extension of v_j^a whenever $a < b$. By Lemmas 2 and 14 of Section 2 of Part I, for all q we have that $G_i(v_j^q/v_j)$ is cyclic and $u_{q+1}(G_i(v_j^{q+1}/v_j)) = G_i(v_j^q/v_j)$ where u_{q+1} denotes the natural homomorphism of $G(K^{q+1}/K)$ onto $G(K^q/K)$. In view of Lemma 8 of Section 2, by induction on q we can fix a generator b_j^q of $G_i(v_j^q/v_j)$ such that $u_{q+1}(b_j^{q+1}) = b_j^q$ for all q. Then by Proposition 16, $b_1^q, \cdots, b_t^q$ generate $G^q = G(K^q/K)$ for each q. Since $G^1 < G^2 < \cdots$ is cofinal in $\pi'(V - W)$, by Lemma 22 of Section 8 of Part I, we can find a group G generated by t generators $a_1, \cdots, a_t$ and a weak parent map f of G onto $\pi'(V - W)$ such that the f image of a_j in G^q is b_j^q for all q and for $j = 1, \cdots, t$. In view of Lemma 2 of Section 2 of Part I and Proposition 4 of Section 9 of Part I, everything now follows from the above Proposition 16.

Remark 7. Remarks 8, 9, and 11 of Section 12 of Part I now apply after replacing reference there to Theorem 2 of Part I by reference to the above Theorems 1 and 2 and replacing the reference there to Theorem 1 of Part I by reference to the above Propositions 15 and 16. Also note that for dimension two, Theorems 1 and 2 of Part I are now subsumed respectively under Proposition 15 and Theorem 1 above.[14]

Remark 8. It is clear that Theorems 1 and 2 together with their proofs can be combined in various ways to get other results, here we shall give three examples of modified assumptions on W leaving the conclusions on $\pi'(V-W)$ to the reader: (1) Assume that the components W_j can be labelled so that

$$\dim|W_j| > 1+\nu(W_j, W; V) \text{ for } j=1,2,\cdots,s,$$

$$\dim|W_j| > 1+\nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t) \quad \text{for } j=s+1, s+2,\cdots,t,$$

and that W_j and W_k have a point in common at which W has a strong normal crossing whenever $j,k \leqq s$ and $j \neq k$. (2) Assume that the components W_j can be labelled so that

$$\dim|W_j| > 1+\nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; V) \text{ for } j=1,\cdots,s$$

and

$$\dim|W_j| > 1+\nu(W_j, W_{s+1} \cup W_{s+2} \cup \cdots \cup W_t; V) \quad \text{for } j=s+1, s+2,\cdots,t.$$

(3) Assume that the components W_j can be labelled so that

$$\dim|W_j| > 1+\nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; V) \text{ for } j=1,\cdots,s;$$

$$\dim|W_j| > 1+\nu(W_j, W_{s+1} \cup W_{s+2} \cup \cdots \cup W_t; V) \quad \text{for } j=s+1, s+2,\cdots,t;$$

and W_j and W_k have a point in common at which $W_{s+1} \cup W_{s+2} \cup \cdots \cup W_t$ has a strong normal crossing whenever $j,k > s$ and $j \neq k$.

10. Applications. Throughout this section k will denote an algebraically closed ground field of characteristic p. Note that the dimension of the complete linear system determined by a curve of degree g^* in a projective plane over k is given by $\binom{g^*+2}{2} - 1 = \frac{1}{2}g^*(g^*+3)$.

THEOREM 3. *Let W be a curve in the projective plane P_2 over k; let $W_1,\cdots,W_t$ be the irreducible components of W; let g^*_j be the degree of W_j*

and let $d = 1$ *in case* $p = 0$ *and* $d =$ *the highest power of* p *which divides* $g^*_1, \cdots, g^*_t$ *in case* $p \neq 0$; *let* $g_j = g^*_j d^{-1}$, *and let* G *be the abelian group on* t *generators* $a_1, \cdots, a_t$ *with the only relation*

$$a_1^{g_1} \cdots a_t^{g_t} = 1.$$

Assume that $\frac{1}{2} g^*_j (g^*_j + 3) > 1 + \nu(W_j, W; P_2)$ *for* $j = 1, \cdots, t$, *and that* W_j *and* W_k *have a point in common at which* W *has a strong normal crossing whenever* $j \neq k$. *Then* G *s a tame fundamental parent group of* $P_2 - W$. *Also* $\pi^*(P_2 - W) = \pi'(P_2 - W)$ *and hence* G *is a reduced fundamental parent group of* $P_2 - W$ *as well.* G *is a direct product of a free abelian group on* $t - 1$ *generators and a cyclic group of order equal to the greatest common divisor of* $g_1, g_2, \cdots, g_t$; *i. e., equal to the greatest common divisor of* $g^*_1, g^*_2, \cdots, g^*_t$ *in case* $p = 0$ *and to the part of this prime to* p *in case* $p \neq 0$.

Proof. Follows from the proof of Theorem 3 of Section 13 Part I after replacing the reference there to Theorem 1 of Part I by reference to Proposition 15 of Section 9 of the present paper.[14]

THEOREM 4. *Let* K *be a two dimensional algebraic function field over* k, *let* V *be a nonsingular projective model of* K/k *and let* W *be an irreducible curve on* V. *Assume that* V *is simply connected and* $\dim |W| > 1 + \nu(W, W; V)$. *Let* K^* *be the compositum of all the fields in* $\Omega'_g(V - W)$. *Then* (i) K^*/K *is cyclic of degree* $\delta(W, V)$; *and* $V - W$ *has as a tame (as well as reduced) fundamental parent group a cyclic group of order* $\delta(W, V)$. *Let* V^* *be a* K^**-normalization of* V *and let* ϕ *be the map of* V^* *onto* V *(so that* $V^* - \phi^{-1}(W)$ *is the "tame universal covering" of* $V - W$*). Then* (ii) $V^* - \phi^{-1}(W)$ *is tamely simply connected. Finally* (iii) *the normalization of* V *in any field between* K *and* K^* *(in particular* V^**) is simply connected.*

Proof. Follows from the proof of Theorem 4 of Section 14 of Part I after replacing the reference there to Theorem 1 and Proposition 6 of Part I, respectively, by reference to Propositions 15 and 14 of Section 9 of the present paper.[14]

PROPOSITION 17. *Let* V^* *be a normal projective surface over* k. *Assume that there exists a rational map* ϕ *of* V^* *onto a nonsingular projective simply connected surface* V *of finite index such that:* (1) ϕ *and* ϕ^{-1} *are both free from fundamental points,* (2) V^*/V *is tamely ramified,* (3) $\Delta(V^*/V)$ *is irreducible, and* (4) $\dim |\Delta(V^*/V)| > 1 + \nu(\Delta(V^*/V), \Delta(V^*/V); V)$. *Then* V^* *is simply connected,* $k(V^*)/k(V)$ *is galois with galois group cyclic of order dividing* $\delta(\Delta(V^*/V), V)$, *and* $V^* - \phi^{-1}(\Delta(V^*/V))$ *is tamely simply connected in case* $[k(V^*) : k(V)] = \delta(\Delta(V^*/V), V)$.

Proof. This is essentially Theorem 4 stated from a covering to the projection instead of the other way around.

PROPOSITION 18. *Let W be an irreducible curve of reduced degree g and degree g^* in the projective plane P_2 over k such that $\frac{1}{2}g^*(g^*+3)>1+\nu(W,W;V)$. Let K^* be the compositum of all the fields in $\Omega'_g(P_2-W)$. Then* (i) *$K^*/k(P_2)$ is cyclic of degree g so that P_2-W has for a tame (as well as reduced) fundamental parent group a cyclic group of order g. Let V^* be a K^*-normalization of P_2 and let ϕ be the rational map of V^* onto P_2. Then* (ii) *$V^*-\phi^{-1}(W)$ is tamely simply connected. Finally* (iii) *the normalization of P_2 in any field between $k(P_2)$ and K^* (in particular V^*) is simply connected.*

Proof. Follows from Theorem 4 in view of Lemma 34 of Section 13 of Part I or alternatively (i) is exactly Theorem 3 for $t=1$ and (ii) and (iii) follow from (i) as in the proof of Theorem 4 of Part I.[14]

THEOREM 5. *Let V^* be a surface in projective 3 dimensional space P_3 over k having an affine equation*

$$X_3^m - f(X_1,X_2) = 0$$

where $W\colon f(X_1,X_2)=0$ is an irreducible curve of degree g^* and reduced degree g (i.e., f is an irreducible polynomial of degree g^*) in the projective plane P_2 over k (with affine coordinates X_1,X_2) such that $\frac{1}{2}g^*(g^*+3)>1+\nu(W,W;P_2)$ and such that m divides g. Then V^* is simply connected. If $m=g$ then $V^*-(f(X_1,X_2)=0\cap V^*)$ is tamely simply connected.

Proof. Follows from the proof of Theorem 5 of Section 14 of Part I after replacing the reference there to Propositions 7 and 8 of Part I, respectively, by above Propositions 17 and 18 and substituting 2 for n.[14]

PROPOSITION 19. *Let V be a nonsingular algebraic surface over the complex ground field such that V has no finite unramified topological coverings (this is so in particular if $\pi_1(V)=1$), let W be a curve on V and let $W_1,\cdots,W_t$ be the irreducible components of W, and by $\gamma\pi_1(V-W)$ denote the factor group of $\pi_1(V-W)$ by the intersection of all subgroups of finite index in $\pi_1(V-W)$.* (1) *If for some labelling of the components W_j, $\dim|W_j|>1+\nu(W_j,W_j\cup W_{j+1}\cup\cdots\cup W_t;V)$ for $j=1,\cdots,t$, then $\gamma\pi_1(V-W)$ is t-step solvable and it is an unrestricted tame fundamental parent group of $V-W$ and its Krull-completion has a dense subgroup generated by t generators.* (2) If $\dim|W_j|>1+\nu(W_j,W;V)$ for $j=1,\cdots,t$ *then $\gamma\pi_1(V-W)$ is t-step nilpotent.* (3) *If* $\dim|W_j|>1+\nu(W_j,W_j,V)$

for $j=1,\cdots,t$ and W_j and W_k have a point in common at which W has a normal crossing whenever $j\neq k$, then $\gamma\pi_1(V-W)$ is abelian.

Proof. Follows from Theorems 1 and 2 of Section 9 in view of the considerations of Section 15 of Part I.

Remark 9. Theorems 3, 4, 5 and Propositions 7, 8, 9 of Part I for $n=2$ are now subsumed, respectively, under above Theorems 3, 4, 5 and Propositions 17, 18, 19. In deducing Theorem 3 we have only used a part of Theorem 1. Using the full force of Theorems 1 and 2 we could have stated the corresponding versions for the projective plane, we have not done this here because elsewhere we shall state stronger conclusions under these circumstances.

Remark 10. It is clear that using Propositions 6 to 10 of Section 7 we can get several explicit corollaries for all the results of this and the previous section. Let us illustrate this by some examples.

Example 1 (*For Theorem* 1). Let V be a nonsingular simply connected algebraic surface over k, let W be a curve on V, and let $W_1,\cdots,W_t$ be the irreducible components of W. Assume that at a common point of distinct irreducible components of W only two irreducible components pass and each has a simple point there. and let j_a be the sum of orders of contact of W_j with the various W_i $(i\neq j)$ at the various points of intersection. Also assume that for each j, the singularities of W_j are all cusps and let $j_1,\cdots,j_{b_j}$ be the orders of these cusps. If

$$\dim|W_j|>1+j_a+\tfrac{1}{2}\sum_{q=1}^{b_j}j_q(j_q+3)$$

for $j=1,\cdots,t$, then $\pi'(V-W)=\pi^*(V-W)$ is generated by t generators and is t-step nilpotent; if in addition W_j and W_i have a point in common at which W has a normal crossing whenever $j\neq i$, then $\pi'(V-W)$ is abelian.

Example 2 (*For Proposition* 18). Let W be an irreducible curve in the projective plane P_2 over k. Let g be the reduced degree of W and let g^* be the degree of W. Assume that any point P of P_2 at most two analytic branches of W have a common tangent and that if two analytic branches of W at P do have a common tangent then each of these branches have a simple point at P; also assume that at any point P of W all analytic branches of W not having a simple point at P necessarily have a cusp at P. Let $u_1,\cdots,u_a$ denote the orders of contact of various pairs of branches having tangents at various points of W $(u_i>1)$; let $v_1,\cdots,v_b$ denote the orders of the cusps of the

various branches of W at various points of W $(v_i > 1)$, and let $w_1, \cdots, w_c$ denote the multiplicities of W at the various singular points of W $(w_i > 1)$. Denote by $\bar{\pi}'(P_2 - W)$ the galois group over $k(P_2)$ of the compositum of all finite extensions of $k(P_2)$ (in a fixed algebraic closure of $k(P_2)$) which are tamely ramified over P_2 and for which the branch locus on P_2 is contained in W. If

$$\tfrac{1}{2}g^*(g^* + 3) > 1 + 3\sum_{i=1}^{a}(u_i - 1) + \sum_{i=1}^{b} v_i + \tfrac{1}{2}\sum_{i=1}^{c} w_i(w_i + 1)$$

then $\bar{\pi}'(P_2 - V)$ is cyclic of order g.

Example 3 (*For Proposition* 18). *Irreducible plane curves of degree* $\leqq 4$. Let W be an irreducible curve of degree g^* in the projective plane P_2 over k and let $\bar{\pi}'(P_2 - W)$ denote the galois group over $k(P_2)$ of the compositum of all finite galois extensions of $k(P_2)$ (in some fixed algebraic closure of $k(P_2)$) which are tamely ramified over P_2 and for which the branch locus on P_2 is contained in W. Here we consider the situation $g^* \leqq 4$.

$g^* = 1$ (*Lines*). $\dim |W| = 2$ and $\nu(W, W; P_2) = 0$. Hence $\bar{\pi}'(P_2 - W) = 1$.

$g^* = 2$ (*Conics*). $\dim |W| = 5$. Again W is nonsingular and therefore $\nu(W, W; P_2) = 0$. Hence $\bar{\pi}'(P_2 - W)$ is cyclic of order 2 or 1 according as $p \neq 2$ or $p = 2$.

$g^* = 3$ (*Cubics*). $\dim |W| = 9$. W can have at most one singularity, because otherwise for a line L joining two singularities of W we would have $i(L \cdot W, P_2) > 3$. Let P be a singularity of W. Then P must be a double point of W, because otherwise for a line L tangent to W at P we would have $i(L \cdot W, P_2) > 3$. If there are two branches of W at P then again for the same reason these two branches cannot have a common tangent and then $\nu(W, W; P, V) = 3$. If there is only one branch of W at P then for the tangent line L to W at P we must have $\nu(L \cdot W; P, P_2) = 3$ and hence P must be a two-fold cusp of W and then $\nu(W, W; P, P_2) = 5$. Thus always $\nu(W, W; P_2) + 1 \leqq 5 + 1 = 6 < 9$. Hence $\bar{\pi}'(P_2 - V)$ is cyclic of order 3 or 1 according as $p \neq 3$ or $p = 3$.

$g^* = 4$ (*Quartics*). We shall show in a later paper that (1) if W does not have three cusps then $\bar{\pi}'(P_2 - W)$ is cyclic of order 4 or 1 according as $p \neq 2$ or $p = 2$; (2) any two quartics with three cusps are projectively equivalent; and (3) if W is an irreducible quartic having three cusps and $p \neq 2, 3$ then $\bar{\pi}'(V - W)$ is a nonabelian group of order 12 and that this

group is the same as the one found by Zariski in the classical case [Zariski 11; Zariski 12, page 164].

Remark 11. In the situations of Theorems 1 and 2 of Section 9; conjecture 2 of Section 16 of Part I has now been verified and in the situation of Theorem 3 of this section, conjecture 1 of Section 16 of Part I has now been verified. Also Remark 14 of Section 17 of Part I now applies to the results of this and the previous section.

PRINCETON UNIVERSITY.

REFERENCES.

[1] S. Abhyankar, "On the ramification of algebraic functions," *American Journal of Mathematics*, vol. 77 (1955), pp. 575-592.

[2] ———, "Local uniformization on algebraic surfaces over ground fields of characteristic $p \neq 0$," *Annals of Mathematics*, vol. 63 (1956), pp. 491-526.

[3] ———, "Coverings of algebraic curves," *American Journal of Mathematics*, vol. 79 (1957), pp. 825-856.

[4] ———, "On the valuations centered in a local domain," *American Journal of Mathematics*, vol. 78 (1956), pp. 321-348.

[5] ———, Ramification theoretic methods in algebraic geometry, Annals of Mathematics Studies, No. 43, Princeton (1959).

[6] C. Chevalley, "On the theory of local rings," *Annals of Mathematics*, vol. 44 (1943), pp. 690-708.

[7] W. L. Chow, "The general connectedness theorem," *American Journal of Mathematics*, vol. 81 (1959), pp. 1033-1074.

[8] I. S. Cohen, "On the structure and ideal theory of complete local domains," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 54-106.

[9] W. Krull, "Der allgemeine Discriminantensatz," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 1-49.

[10] ———, Idealtheorie, New York (1948).

[11] O. Zariski, "On the problem of existence of algebraic functions of two variables possessing a given branch curve," *American Journal of Mathematics*, vol. 51 (1939), pp. 305-328.

[12] ———, Algebraic surfaces, New York (1948).

[13] ———, Theory and applications of holomorphic functions on algebraic varieties over arbitrary ground fields, New York (1951).

[14] ———, "A simplified proof for the resolution of singularities of an algebraic surface," *Annals of Mathematics*, vol. 43 (1942), pp. 583-593.

[15] ———, "Foundations of a general theory of birational correspondences," *Transactions of the American Mathematical Society*, vol. 53 (1943), pp. 490-542.

[16] ———, "On the purity of the branch locus of algebraic functions," *Proceedings of National Academy of Sciences, U. S. A.*, vol. 44 (1958), pp. 791-796.

TAME COVERINGS AND FUNDAMENTAL GROUPS OF ALGEBRAIC VARIETIES.*

Part III: Some Other Sets of Conditions for the Fundamental Group to be Abelian.

By SHREERAM ABHYANKAR.

Dedicated to my teacher Professor Oscar Zariski on his sixtieth birthday.

Introduction. Let K be an n dimensional algebraic function field over an algebraically closed ground field k of characteristic p, and let V be a simply connected nonsingular projective model of K/k, let W be a pure $(n-1)$-dimensional subvariety of W and let $W_1, \cdots, W_t$ be the irreducible components of W. Denote by $\pi'(V-W)$ the group tower of the galois groups over $k(V)$ of all the finite galois extensions of $k(V)$ (in some fixed algebraic closure of $k(V)$) which are tamely ramified over V and for which the branch locus over V is contained in W.

In Theorem 2 of Section 12 of Part I[1,2] we proved that if (i) W has only strong normal crossings (ii) $\dim |W_j| > 1$ for each j and (iii) the components W_j are pairwise connected, then $\pi'(V-W)$ is abelian with t generators; also in Theorem 1 of Section 9 of Part II we proved that if (i) $n=2$, (ii)[3] $\dim |W_j| > 1 + \nu(W_j, W, V)$ for each j and (iii) W_j and W_k have a point in common at which W has a normal crossing whenever $j \neq k$; then $\pi'(V-W)$ is abelian with t generators. In this paper we want to show that conditions (iii) above can be replaced by the condition that for each j

* Received February 18, 1959.

[1] "Part I: Branch loci with normal crossings; Applications: Theorems of Picard and Zariski," and "Part II: Branch curves with higher singularities," published in respectively in volumes 81 (1959) and 82 (1960) of this Journal; these will be referred to as "Part I" and "Part II" respectively. Notations, conventions and definitions given in Parts I and II will be followed and some more will be introduced in Section 1.

[2] Also see the correction given in Remark 6 of Section 8 of Part II.

[3] $\nu(W_j, W; V)$ is a certain measure of the singularities of W_j relative to W; for definition see Section 6 of Part II.

and k some nonzero integral multiple of W_j is linearly equivalent to some integral multiple of W_k (Theorem 1 of Section 3), and then in the second mentioned result condition (ii) can be replaced by the weaker condition that for some labelling of the components W_j,

$$\dim |W_j| > 1 + \nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; V)$$

for each j; this will be done by using the method of "removing tame ramification through cyclic compositum." (Proposition 1 of Section 2). Consequently in case V is the n dimensional projective space over k then we can drop conditions (iii) altogether, thus yielding a further generalization of Zariski's theorem [Part I, Theorem 3 of Section 13; Part II, Theorem 3 of Section 10] (Theorem 2 of Section 4).

1. Notations and conventions. Let v be a real discrete valuation of a field K, let p be the characteristic of the residue field of v, let K^* be a finite separable extension of K, and let v^* be a K^*-extension of v. In [A1, Section 1; A2, Section 2; Part I, Section 2][4] we developed various notions of ramification and galois theories of quotient rings on normal algebraic varieties; the corresponding notions for real discrete valuations are well known and they also follow from the quoted reference since the only special properties of a normal local domain when it is the quotient ring on an algebraic variety which were used were that it is noetherian and its completion is also a normal local domain. Consequently we may and we shall use the notions, notations and results given in the above refrence also for real discrete valuations. Then if $\bar{R}_{v^*}$, $\bar{R}_v$ are the completion of R_{v^*} and R_v respectively and if E^* and E are the quotient fields $\bar{R}_{v^*}$ and $\bar{R}_v$ respectively then

$$\begin{aligned}
&d(v^*: v) = \text{degree of } v^* \text{ over } v\\
&= d(R_{v^*}: R_v) = \text{degree of } R_{v^*} \text{ over } R_v\\
&= d(v^*: K) = \text{degree of } v^* \text{ over } K\\
&= [E^*: E].
\end{aligned}$$

$$\begin{aligned}
&g(v^*: v) = \text{separable residue degree of } v^* \text{ over } v\\
&= g(R_{v^*}: R_v) = \text{separable residue degree of } R_{v^*} \text{ over } R_v\\
&= g(v^*: K) = \text{separable residue degree of } v^* \text{ over } K\\
&= [(R_{v^*}/M_{v^*}) : (R_v/M_v)]_s.
\end{aligned}$$

$$\begin{aligned}
&i(v^*: v) = \text{inseparable residue degree of } v^* \text{ over } v = \text{etc.}\\
&= [(R_{v^*}/M_{v^*}) : (R_v/M_v)]_i \text{ in case } p \neq 0 \text{ and } 1 \text{ in case } p = 0.
\end{aligned}$$

[4] Numbers in square brackets refer to the references at the end of the paper.

$r(v^*:v)$ = ramification index of v^* over v = etc.
$= d(v^*:v)g(v^*:v)^{-1}$.

$\bar{r}(v^*:v)$ = reduced ramification index of v^* over v = etc.
$= r(v^*:v)i(v^*:v)^{-1}$.

Now it follows from well known results [for instance C, Theorem 23] that the reduced ramification index of v^* over v is the index of the value group of v in the value group of v^*. Hence in particular if $v^*_1, \cdots, v^*_t$ are distinct extensions of v to K^* then $\sum_{j=1}^{t} \bar{r}(v^*_j:v)g(v^*_j:v)i(v^*_j:v) = [K^*:K]$. Note that if R_v is the quotient ring of an irreducible one codimensional subvariety W of an algebraic variety then the ramification and galois theoretic notions for v coincide with those for W.

For a polynomial $f = f(X_1, \cdots, X_n)$ we denote by $\Delta_{X_1}f = \Delta_{X_1}f(X_1, \cdots, X_n)$ the discriminant of f when f considered as a polynomial in X_1 whose coefficients are polynomials in $X_2, \cdots, X_n$; when the reference is clear from the context, the subscript X_1 may be dropped.

2. Removing tame ramification through cyclic compositums. In [A1, A3] and Parts I and II we have several times used a method which can be called "removing the tame ramification of a real discrete valuation through compositums with a cyclic extension," we give this in a general form in Proposition 1 below and it then subsumes Lemma 6 of [A1] and Proposition 8 of [A3].

LEMMA 1. *Let R be either the quotient ring of an irreducible subvariety of a normal algebraic variety or the valuation ring of a real discrete valuation. Let K be the quotient field of R, let L and K^* be finite separable extensions of K, let L^* be a compositum of L and K^*, let S^* be a local ring in L^* lying above R, let $R^* = K^* \cap S^*$ and $S = L \cap S^*$. Assume that S is unramified over R. Then S^* is unramified over R^* and $r(S^*:S) = r(R^*:R)$, $\bar{r}(S^*:S) = \bar{r}(R^*:R)$, $i(S^*:S) = i(R^*:R)$, and $g(S^*:S) \leqq g(R^*:R)$.*

Proof. Since the situation remains parallel if we pass to the completions of R, S, R^*, S^* [Section 2 of A2], we may assume that these local domains are complete to begin with. Let D, D^*, E, E^* be the residue fields of R, R^*, S, S^* respectively. Then our assumption implies that $[E:D] = [E:D]_s = [L:K]$. Since E/D is separable, there exists a in E such that $E = D(a)$. Fix an element A in S belonging to the residue class a. Let $F(X)$ be the minimal monic polynomial of A over K. Then all the coefficients of F are in R. Let

$f(X)$ be the monic polynomial obtained by reducing the coefficients of $F(X)$ modulo the maximal ideal in R. Let e denote the degree of F and hence also the degree of f. Now we must have $f(a)=0$ and hence $e \geqq [E:D]$. However also $e=[K(A):K] \leqq [L:K]$. Since $[E:D]=[L:K]$, we have $e=[E:D]=[L:K]$ and hence $K(A)=L$. Consequently, $K^*(A)=L^*$. Let $G(X)$ be the minimal monic polynomial A over K^*. Since R^* is normal, $F=GH$ where G and H are monic polynomials with coefficients in R^*. Hence $\Delta(G)$ divides $\Delta(F)$ in R^*. Now $\Delta(F)$ belongs to the residue class (modulo the maximal ideal in R) $\Delta(f)$. Since f is a separable polynomial, $\Delta(f) \neq 0$ and hence $\Delta(F)$ is a unit in R and hence a unit in R^*. Therefore $\Delta(G)$ is a unit in R^* and consequently [see K] S^* is unramified over R^*, i.e., $r(S^*:R^*)=\bar{r}(S^*:R^*)=i(S^*:R^*)=1$.

Now $r(S^*:S)r(S:R)=r(S^*:R)=r(S^*:R^*)r(R^*:R)$ and hence $r(S^*:S)=r(R^*:R)$, and similarly $\bar{r}(S^*:S)=\bar{r}(R^*:R)$ and $i(S^*:S)=i(R^*:R)$. Next, $g(S^*:S)r(S^*:S)=[L^*:L]$ and $g(R^*:R)r(R^*:R)=[K^*:K]$; since L^* is the compositum of K^* and L, any set of K-generators of the K-vector space K^* is also a set of L-generators of the L-vector space L^* and hence $[L^*:L] \leqq [K^*:K]$ and therefore in view of the equality $r(S^*:S)=r(R^*:R)$ we can conclude that $g(S^*:S) \leqq g(R^*:R)$.

Remark 1. The inequality $g(S^*:S) \leqq g(R^*:R)$ is in general false if we do not assume that S is unramified over R. This can be seen from the following example. Let k be an algebraically closed field of characteristic p, let u and x be independent variables over k, let n be an integer greater than 1 such that n is prime to p in case $p \neq 0$. Let $K=k(u)((x))$, $L^*=K(x^{1/n}, u^{1/n})$, $K^*=K(x^{1/n})$, $L=K(x^{1/n}u^{1/n})$, $R=k(u)[[x]]$. Then we must have

$$S^*=k(u^{1/n})[[x^{1/n}]], \quad R^*=k(u)[[x^{1/n}]], \quad S=k(u)[[(ux)^{1/n}]].$$

Consequently $g(R^*:R)=1$ and $g(S^*:S)=n$.

LEMMA 2. *Let v be a real discrete valuation of a field K, let p be the characteristic of the residue field of v, let $n_1, \cdots, n_t$ be positive integers which are prime to p in case $p \neq 0$, let $x_1, \cdots, x_t$ be elements of K having v-value zero, and let L be a field generated over K by a certain number of roots of the polynomial $\prod_{j=1}^{t}(X^{n_j}-x_j)$. Then v is unramified in L. Now let K^* be a finite separable extension of K, let L^* be a composition of L and K^*, let w^* be an extension of v to L^*, let v^* be the restriction of w^* to K^* and let w be the restriction of w^* to L. Then $r(w^*:w)=r(v^*:v)$, $\bar{r}(w^*:w)=\bar{r}(v^*:v)$, $i(w^*:w)=i(v^*:v)$.*

Proof.[5] If T is a finite separable extension of K and T^* is a finite separable extension of T, then v is unramified in T if and only if $r(u:v)=1$ for every T-extension u of v; also if u is a T-extension of v and if u^* is a T^*-extension of u then $r(u^*:v)=r(u^*:v)r(u:v)$, $u(x_j)=0$ for all j. Therefore it is clear that the first assertion will be proved if we show that v is unramified in $K(z)$ where z is a root of $f=f(X)=X^{n_1}-x_1$; let $g(X)$ be the minimal polynomial of z over K, since R_v is integrally closed we must have $f=gh$ where g and h are monic polynomials with coefficients in R_v and hence $\Delta(g)$ divides $\Delta(f)$ in R_v; now $\Delta(f)$ equals a power of n_1 times a power of x_1, x_1 is a unit in R_v and n_1 is not divisible by p if $p\neq 0$ and hence n_1 is also a unit in R_v, therefore $\Delta(f)$ and consequently $\Delta(g)$ is a unit in R_v, from this we conclude that v is unramified in $K[z]$ (see [K]); this completes the proof of the first assertion. The second assertion now follows from Lemma 1.

LEMMA 3. *Let v be a real discrete valuation of a field K, let p be the characteristic of the residue field of v, let x be a nonzero element of K, let $q=v(x)$, let n be a positive integer which is prime to p in case $p\neq 0$, let ν be the greatest common divisor of n and q,*[6] *let $b=n/\nu$ and K^* be a field generated over K by one or more roots of the polynomial X^n-x over K and let v^* be an extension of v to K^*. Then $r(v^*:v)=\bar{r}(v^*:v)=b$ and $i(v^*:v)=1$.*

Proof.[5] Fix y in K with $v(y)=1$ and let $A=x/y^q$. Then $v(A)=0$, and hence $\alpha\neq 0$. Let L be a field generated over K by all the roots of the polynomial $(X^n-A)(X^n-1)$, let L^* be a compositum of K^* and L, let w^* be an extension of v^* to L^* and let w be the L-restriction of w^*. Then by Lemma 2, $r(v^*:v)=r(w^*:w)$, $\bar{r}(v^*:v)=\bar{r}(w^*:w)$ and $w(y)=1$. Therefore it is enough to show that $r(w^*:w)=\bar{r}(w^*:w)=b$. By our assumption, L^* contains a root z of X^n-x and L^* is generated over L by a certain number of roots of X^n-r. Let B be a root of X^n-A in L and let $\xi=z/B$. Then $\xi\in L^*$ and $\xi^n=y^q$. Let t be any root of X^n-x (in some field extension of L^*), then $t^n=x=z^n$ and hence t/z is a root of X^n-1; therefore $t/z\in L^*$ and hence $t\in L^*$. Therefore $L^*=L(z)=L(\xi)$. Let $a=q/\nu$. Now X^n-1 has all its roots in L and n is prime to p in case $p\neq 0$ and hence n is prime to the characteristic of L in case this is different from zero, therefore L contains a primitive n-th root h of unity. Now $(\xi^b/y^a)^\nu=\xi^{b\nu}/y^{a\nu}=\xi^n/y^q=1$

[5] Our assumption implies that all the extensions considered in the statement of the lemma and all the extensions to be considered in the proof are separable and hence we can use the methods of ramification and galois theories.

[6] Note that if $q=0$ then $\nu=n$ and $b=1$.

and $b\nu = n$, therefore $\xi^b/y^a = h^{sb}$ where s is an integer. Let $\eta = \xi h^{-s}$. Then $L^* = L(\eta)$ and $\eta^b = y^a$. If $a = 0$ then $b = 1$ and there is nothing to prove, so now assume that $a \neq 0$. Then a and b are coprime nonzero integers and hence we can find integers α and β such that $a\alpha + b\beta = 1$. Let $\zeta = \eta^\alpha y^\beta$. Then $\zeta^a = \eta^{\alpha a} y^{\beta a} = \eta^{\alpha a}(y^a)^\beta = \eta^{\alpha a}(\eta^b)^\beta = \eta$ and hence $L^* = L(\zeta)$. Also $\zeta^b = \eta^{\alpha b} y^{\beta b} = (\eta^b)^\alpha y^{b} = (y^a)^\alpha y^{b\beta} = y$. Therefore $\bar{r}(w^*: w) = w(y)\bar{r}(w^*: w) = w^*(y) = w^*(\zeta^b) = bw^*(\zeta)$ and hence $\bar{r}(w^*: w) \geqq b \geqq [L^*: L]$. Therefore $\bar{r}(w^*: w) = b = [L^*: L]$. Hence $r(w^*: w) = \bar{r}(w^*: w) = b$.

PROPOSITION 1. *Let v be a real discrete valuation of a field K, let p be the characteristic of the residue field of v, let K^* be a finite separable extension of K, let v^* be an extension of v to K, let $x, x_1, x_2, \cdots, x_t$ $(t \geqq 0)$ be nonzero elements of K such that $v(x_j) = 0$ for $j = 1, \cdots, t$; and let $n = n_0, n_1, \cdots, n_t$ be positive integers which are not divisible by p in case $p \neq 0$. Let L be a field generated over K by one or more roots of the polynomial $X^n - x$ and a certain number of roots of the polynomial $\prod_{j=1}^{t}(X^{n_j} - x_j)$, let L^* be a compositum of K^* and L, let w^* be an extension of v^* to L^* and let w be the restriction of w^* to L. Let $q = v(x)$ and let $a = \bar{r}(v^*: v)$. Let ν be the greatest common divisor of q and n.*[6] *Let $b = n/\nu$. Let d be the greatest common divisor of a and b. Let $e = a/d$. Then* (i) *$\bar{r}(w^*: w) = e$ and $i(w^*: w) = i(v^*: v)$. Furthermore* (ii) *if $q = 1$ and a divides n then $\bar{r}(w^*: w) = 1$;* (iii) *if $r(v^*: v) \not\equiv 0 \pmod p$ in case $p \neq 0$ and always in case $p = 0$ we can choose x and n such that $q = 1$ and a divides n, and then w^* is unramified over w; and* (iv) *if $q = 1$, if a divides n, if $i(v^*: v) = 1$, if T is a finite separable extension of L; if T^* is the compositum of T and K^*, if u^* is a T^*-extension of v^* and if u is the T-restriction of u^*, then u^* is unramified over u.*

Proof.[5] (ii) and (iii) follow from (i), and (iv) follows from (iii) in view of Lemma 1; hence it is enough to prove (i).

Let K_1 be a field generated over K by all the roots of the polynomial $\prod_{j=1}^{t}(X^{n_j} - x_j)$; let K^*_1, L_1, L^*_1 be the compositums of K_1 respectively with K^*, L, L^*; let w^*_1 be an extension of w^* to L^*_1 and let v_1, v^*_1, w_1 be the restrictions of w^*_1 respectively to K_1, K^*_1, L_1. Then by Lemma 2 we have that

$$\bar{r}(v^*_1: v_1) = \bar{r}(v^*: v) = a, \qquad i(v^*_1: v_1) = i(v^*: v), \bar{r}(w^*_1: w) = \bar{r}(w^*: w), i(w^*_1: w_1) = i(w^*: w); \text{ and } v_1(x) = v(x) = q \cdots \text{(I)}$$

Now $n = b\nu$ and $q = c\nu$ where b and c are coprime nonzero integers and

$a = ed$ and $b = fd$ where e and f are coprime nonzero integers. Applying Lemma 3 to L_1/K_1 we have that

$$\bar{r}(w_1 : v_1) = b \text{ and } i(w_1 : v_1) = 1 \cdots (\mathrm{II})$$

Now $aq = (ec)(d\nu)$ and $n = (f)(d\nu)$; since b and c are coprime, f and c must also be coprime; also f and e are coprime and hence f and ec are coprime; therefore $d\nu$ is the greatest common divisor of aq and n, and $n = (f)(d\nu)$. Also by (I), $v^*_1(x) = \bar{r}(v^*_1 : v_1)v_1(x) = aq$. Hence applying Lemma 3 to L^*_1/K^*_1 we get

$$\bar{r}(w^*_1 : v^*_1) = f \text{ and } i(w^*_1 : v^*_1) = 1 \cdots (\mathrm{III})$$

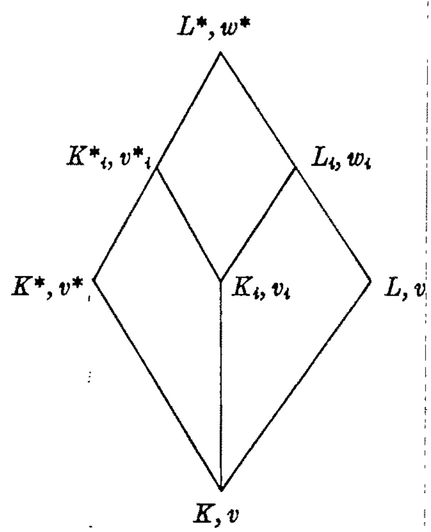
Now $\bar{r}(w^*_1 : v^*_1)\bar{r}(v^*_1 : v_1) = \bar{r}(w^*_1 : v_1) = \bar{r}(w^*_1 : w_1)\bar{r}(w_1 : v_1)$ and hence by (I, II, III) we get $fa = \bar{r}(w^*_1 : w_1)b$ and substituting ed for a and fd for b this gives us $\bar{r}(w^*_1 : w_1) = e$ and hence by (I) : $\bar{r}(w^* : w) = e$. Again $i(w^*_1 : v^*_1)i(v^*_1 : v_1) = i(w^*_1 : v_1) = i(w^*_1 : w_1)i(w_1 : v_1)$ and hence by (I, II, III) we have: $i(w^* : w) = i(v^* : v)$.

PROPOSITION 2. *Let K be an n dimensional algebraic function field over an algebraically closed ground field k, let V be a nonsingular projective model of K/k, let W be a pure $n-1$ dimensional subvariety of K and let $W_1, \cdots, W_t$ be the irreducible components of W. Assume that the following conditions are satisfied:* (i) *$\pi'(V - W_2 - W_3 - \cdots - W_t)$ is abelian*; (ii) *$Q(W_1, V)$ does not split (i. e. there is a unique local ring lying above it) in any finite separable extension of K which is tamely ramified over V and for which the branch locus over V is contained in W*; (iii) *some nonzero integral multiple of W_1 is linearly equivalent to some integral multiple of W_2. Then $\pi'(V - W)$ is abelian.*

Proof.[7] We have to show that if K^*/K is a galois extension such that K^*/V is tamely ramified and $\Delta(K^*/V) \subset W$ then $G(K^*/K)$ is abelian. Assumption (iii) means that there exists y in K such that $(y) = \alpha W_1 - \beta W_2$ where α and β are nonzero integers. Let v be the real discrete valuation of K having $Q(W_1, V)$ as its valuation ring and let v^* be the unique K^*-extension of v. Let $a = \bar{r}(v^* : v) = r(v^* : v)$. Let $\alpha_1 = 1$ in case $p = 0$ and $\alpha_1 =$ the highest power of p which divides α in case $p \neq 0$. Let $\alpha_2 = \alpha/\alpha_1$ and let $m = a\alpha_2$. Then m is prime to p in case $p \neq 0$. Let L be the field generated over K by a root x of the polynomial $X^m - y$. Since k is algebraically closed, L/K is a galois extension, $G(L/K)$ is cyclic and its order is prime to p in case $p \neq 0$ and hence L/V is tamely ramified. Since $(y) = \alpha W_1 - \beta W_2$,

[7] All extensions of K are to be taken in some fixed algebraic closure of K.

Lemma 2 tells us that $\Delta(L/V) \subset W_1 \cup W_2 \subset W$. Let L^* be the compositum of L and K^*. Then L^*/K is galois and by Lemma 12 of Part I, L^*/V is tamely ramified and $\Delta(L^*/V) \subset W$. Hence by assumption (ii), there is only one L^*-extension w^* of v and it is clear that v^* is the K^*-restriction of w^*. Let K_i be the inertia field of w^*/v, let K^*_i and L_i be the compositums of K_i with K^* and L respectively, let v_i, v^*_i, w_i be the restrictions of w^* to K_i, K^*_i, L_i respectively. Then by Lemmas 1, 2, 13 and 17 of Part I we get (1) K_i/K is galois, (2) $g(w^*: v_i) = 1$, (3) $r(v_i: v) = 1$, (4) $\Delta(K_i/K) \subset W_2 \cup \cdots \cup W_t$.



Also since L^*/V is tamely ramified, so is K_i/V and hence in view of (4) assumption (i) tells us that $G(K_i/K)$ is abelian, and since L/K is cyclic we conclude that (5) L_i/K is galois and $G(L_i/K)$ is abelian. From (3) and Lemma I we get (6) $\bar{r}(v^*_i: v_i) = a$. Now it is clear that (7) L^* is the compositum of K^*_i and L_i, $L_i = K_i(x)$, and $x^m = y$. Also $(y) = \alpha W_1 - \beta W_2$ implies that $v(y) = \alpha$ and hence by (3) we get (8) $v_i(y) = \alpha = \alpha_1\alpha_2$. Let μ be the greatest common divisor of m and α, and let $b = m/\mu$. Now a is not divisible by p in case $p \neq 0$ and hence always a and α_1 are coprime; since $m = a\alpha_2$ we can conclude that $\mu = \alpha_2$ and hence $b = a$. In view of (6, 7, 8), Proposition 1(i) applied to the top quadrilateral in the diagram now yields (9) $r(w^*: w_i) = 1$. Also (2) or alternatively [A4, Proposition 1.49] tell us that (10) $g(w^*: w_i) = 1$. Since w^* is the only L^*-extension of w_i, (9) and (10) now yield that $L^* = L_i$ and hence by (5), $G(L^*/K)$ is abelian. Therefore $G(K^*/K)$ is abelian.

3. Main result. Throughout this section K will denote an n-dimensional algebraic function field over an algebraically closed ground field k of characteristic p, V will denote a simply connected nonsingular projective model of K/k, W will denote a pure $(n-1)$-dimensional subvariety of V and $W_1, \cdots, W_t$ will denote the irreducible components of W.

PROPOSITION 3. *Assume that for any j and k some nonzero integral multiple of W_j is linearly equivalent to some integral multiple of W_k, and that the components W_j can be labelled such that for $j = 1, \cdots, t$, $Q(W_j, V)$ does not split in any finite separable extension K^* of K for which K^*/V is tamely ramified and $\Delta(K^*/V) \subset W_j \cup W_{j+1} \cup \cdots \cup W_t$. Then $\pi'(V - W)$ is abelian and $\pi^*(V - W) = \pi'(V - W)$.*

Proof. That $\pi'(V - W)$ is abelian follows for $t = 1$ as in the proof of Theorem 4 of Section 14 of Part I and from this the general case follows by induction in view of Proposition 2 above. Now assume that $p \neq 0$ and let K^* be a galois extension of K such that K^*/V is tamely ramified and $\Delta(K^*/V) \subset W$. Then K^*/V is abelian and hence for each j, the inertia groups over K of any local ring in K^* lying above $Q(W_j, V)$ coincide with each other, let G_j be this common group, let G be the subgroup of $G(K^*/K)$ generated by $G_1, \cdots, G_t$ and let L be the fixed field of G. Then by Lemmas 13 and 17 of Part I, L/V is unramified and hence $L = K$. Now the order of each subgroup G_j is prime to p and hence the order of $G(K^*/K)$ is also prime to p. Therefore $\pi^*(V - W) = \pi'(V - W)$.

THEOREM 1. *Assume that for each j and k some nonzero integral multiple of W_j is linearly equivalent to some integral multiple of W_k, and either that W has only strong normal crossings and $\dim |W_j| > 1$ for each j or that $n = 2$ and for some labelling of the components W_j we have $\dim |W_j| > 1 + \nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; V)$ for each j. Then $\pi'(V - W)$ is abelian and is generated by t-generators.*

Proof. In view of Proposition 3, that $\pi'(V - W)$ is abelian follows from Proposition 6 of Section 11 of Part I and Proposition 14 of Section 9 of Part II respectively. That $\pi'(V - W)$ is generated by t generators has been proved in Theorem 2 of Part I[2] and Theorem 2 of Part II.

4. Applications. As in the proofs of Theorem 3 of Section 13 of Part I[2] and Theorem 3 of Section 10 of Part II, Theorem 1 of the previous section now gives the following theorem which subsumes these results of Parts I and II. This is the theorem of which we spoke of in Remark 9 of Section 10 of Part II.

THEOREM 2. *Let P_n be the n dimensional projective space $(n > 1)$ over an algebraically closed ground field k of characteristic p, let W be a hypersurface in P_n with irreducible components $W_1, \cdots, W_t$, let g^*_j be the degree of W_j; let $d = 1$ in case $p = 0$ and $d =$ the highest power of p which divides $g^*_1, \cdots, g^*_t$ in case $p \neq 0$; let $g_j = g^*_j d^{-1}$, and let G be the abelian group generated by t generators $a_1, \cdots, a_t$ with the only relation*

$$a_1^{g_1} \cdots a_t^{g_t} = 1.$$

Assume either that W has only strong normal crossings or that $n = 2$ and the components W_j can be labelled so that

$$\tfrac{1}{2} g^*_j(g^*_j + 3) > 1 + \nu(W_j, W_j \cup W_{j+1} \cup \cdots \cup W_t; P_2).$$

Then G is a tame fundamental parent group of $V - W$. Also $\pi^*(V - W) = \pi'(V - W)$ and hence G is a reduced fundamental parent group of $V - W$ as well. G is a direct product of a free abelian group on $t - 1$ generators and a cyclic group of order equal to the greatest common divisor of $g_1, \cdots, g_t$; i.e. equal to the greatest common divisor of $g^*_1, \cdots, g^*_t$ in case $p = 0$ and to the part of this prime to p in case $p \neq 0$.

Remark 2. Results similar to Proposition 3 with regard to Theorem 2 of Section 12 of Part I, Theorems 1, 2 of Section 9 of Part II and Theorem 2 of this section clearly hold. Also all the remarks made in Parts I and II concerning Theorems 2 and 3 of Part I and Theorems 1, 2 and Part II now are valid in the situations of Theorems 1 and 2 of this paper. Finally it is clear that in the classical case we now have the form of Proposition 19 of Section 10 of Part II corresponding to Theorems 1 and 2 above.

Remark 3. Let the notation be as in Definition 7 of Section 6 of Part II and assume that $A = \mathrm{Rad}_R A$. Then instead of treating a 2-fold normal crossing in a way similar to an s-fold ordinary point $(s \geqq 2)$ one might conceivably have thought of defining "$\nu(A, A; R, R) = 0$ if A has a normal crossing at R." This would have been inappropriate, for otherwise the above Theorem 2 would have been false. To show this let us further consider the example given in Remark 5 of Section 8 of Part II. We shall use the notation of that Remark. Let v be the real discrete valuation of K/k having center W_1 on V. Then as shown, v splits into two valuations in K^*, let these be v' and v^*. Let $a = yz - x$ and $b = yz + x$. Then $v(ab) = v(f) = 1$, and a and b generate distinct prime ideals in R_1 and hence after a suitable labelling of v', v^* we must have $v'(a) = 1$, $v'(b) = 0$, $v^*(a) = 0$, $v^*(b) = 1$. Let W' and W^* be the irreducible components of $\phi^{-1}(W_1)$ corresponding to v', and v^*

respectively. Now the K^*/K norm of a is f and hence the part of the divisor of a on V^* at finite distance equals W'.[8] Let n be any integer greater than 1 such that n is prime to p in case $p \neq 0$. Let $K_1 = K^*(a^{1/n})$. Then K_1/V is tamely ramified and $\Delta(K_1/V) \subset W$. However v' is ramified in K_1 while v^* is not and consequently K_1/K is not galois and hence $\pi'(V - W)$ cannot be abelian, i.e., Theorem 2 does not apply to $V - W$.

Next, one can easily verify the following: (1) the only singularity of W_1 is the 2-fold normal crossing at $P: X = Y = 0$ and neither L nor L_∞ pass through this; (2) at $P': (X = 0, Y = 1)$, L and W_1 have a 2-fold contact and L_∞ does not go through P': (3) at the point P^* at infinity in the direction $Y = 0$, W_1 and L_∞ have a 3-fold contact and L has a normal crossing with W_1. Therefore by the results of Section 7 of Part II we get $\nu(W_1, W; P, V) = 3$, $\nu(W_1, W; P', V) = 2$ and $\nu(W_1, W; P^*, V) = 3$. Hence $\nu(W_1, W; V) = 8$. Also $\dim |W_1| = 9$ and $9 \not> 1 + 8$. This accounts for the nonapplicability of Theorem 2. However had we set $\nu(W_1, W; P, V) = 0$ at the 2-fold normal crossing P then $\nu(W_1, W; V)$ would have been equal to 5 and we would have had $9 > 1 + 5$.

Remark 4. Let P^2 be a projective plane over an algebraically closed ground field k of characteristic p, let W be a curve on P^2, and let $W_1, \cdots, W_t$ be the irreducible components of W. A weaker form of Proposition 3 is this: (A) If for $j = 1, \cdots, t, Q(W_j, P^2)$ does not split in any member of $\Omega'(P^2 - W)$, then $\pi'(P^2 - W)$ is abelian. The corresponding form of Theorem 1 is this: (B) If $\dim W_j > 1 + \nu(W_j, W; P^2)$ for $j = 1, \cdots, t$, then $\pi'(P^2 - W)$ is abelian. Here we shall show that the converse of (A) does not hold, i.e., some W_j can split in some member of $\Omega'(P^2 - W)$ and $\pi'(P^2 - W)$ may still be abelian. The same examples will also show that there exists a tamely ramified covering $f: V \to P^2$, and two distinct lines L_1, L_2 such that $\pi'(V - L^* - f^{-1}(L_2)) = 1$ where L^* is an irreducible component of $f^{-1}(L_1)$.

Let P^3 be a projective three space over k with affine coordinates X, Y, Z; let V be a nonsingular surface in P^3; consider P^2 to be the (X, Y)-plane; and let f be the projection of V onto P^2 along the Z-axis. *Example* 1. $(p \neq 2)$. $t = 3$; $V: Z^2 + XY - 1 = 0$; $W_1: XY - 1 = 0$; $W_2: X = 0$; $W_3 =$ the line at infinity. Then $\Delta(V/P^2) = W_1 \subset W$. $Z^2 + XY - 1 \equiv (Z + 1)(Z - 1)$ $(\mathrm{mod}\, X)$ and hence $f^{-1}(W_2)$ has two irreducible components W^* and W'. Now $\nu(W_1, W; P^2) = 2$ and $\dim |W_1| = 5$. Hence by Theorem 1, $\pi'(P^2 - W)$ is abelian. Suppose if possible that $\pi'(V - W^* - f^{-1}(W_3)) \neq 1$. Then there exists a finite separable extension K of $k(V)$ other than $k(V)$ such that K/V

[8] One can also easily show that $\phi^{-1}(L_\infty)$ is irreducible and $(a) = W_1 - \phi^{-1}(L_\infty)$.

is tamely ramified and $\Delta(K/V) \subset W^* \cup f^{-2}(W_3)$. Then K/P^2 is tamely ramified and $\Delta(K/P^2) \subset W$. Since $\pi'(P^2 - W)$ is abelian, K/P^2 is galois and since W^* and W' are components of $f^{-1}(W_2)$, we must have $\Delta(K/V) \subset f^{-1}(W_3)$ and hence $\Delta(K/P^2) \subset W_1 \cup W_3$. From this, in view of [Part I, Section 13] we conclude that $f^{-1}(W_1)$ must be ramified in K. This is a contradiction. Therefore $\pi'(W - W^* - f^{-1}(W_3)) = 1$. If k is the field of complex numbers then $\pi_1(V - W^* - f^{-1}(W_3)) = 1$. This can be seen by noting that $V - W^* - f^{-1}(W_3)$ is biregularly equivalent to an immediate quadratic transform of a complex affine plane.

Example 2. $(p \neq 3)$. $t = 3$; $V: Z^3 - Y^3 - X(X+a)(X+b) = 0$ where a and b are distinct nonzero elements of k; $W_1: Y^3 + X(X+a)(X+b) = 0$; $W_2: X = 0$; $W_3 =$ line at infinity. Then $\nu(W_1, W; P^2) = 3$, $\dim |W_1| = 9$, and

$$Z^3 - Y^3 - X(X+a)(X+b) \equiv (Z - hY)(Z - h^2Y)(Z - Y) \pmod{X}$$

where h is a primitive cube root of 1, etc.

PRINCETON UNIVERSITY.

REFERENCES.

[A1] S. Abhyankar, "On the ramification of algebraic functions," *American Journal of Mathematics*, vol. 77 (1955), pp. 575-592.

[A2] ———, "Local uniformization on algebraic surfaces over ground fields of characteristic $p \neq 0$," *Annals of Mathematics*, vol. 63 (1956), pp. 491-526.

[A3] ———, "Coverings of algebraic curves," *American Journal of Mathematics*, vol. 79 (1957), pp. 825-856.

[A4] ———, Ramification theoretic methods in algebraic geometry, Princeton (1959).

[C] I. S. Cohen, "On the structure and ideal theory of complete local domains," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 54-106.

[K] W. Krull, "Der allgemeine Discriminantensatz," *Mathematische Zeitschrift*, vol. 45 (1959), pp. 1-49.

ON A CONJECTURE OF LITTLEWOOD AND IDEMPOTENT MEASURES.*

By Paul J. Cohen.

Introduction. In this paper we shall be concerned with two problems. The first is a problem of Littlewood [1] in classical Fourier analysis concerning a lower bound for the L^1 norm of certain exponential sums. The second is the problem of determining all the idempotent measures on a locally compact abelian group. This second problem we shall solve entirely and thus complete a line of investigation begun by Helson [3] and Rudin [5]. The problem of idempotent measures is related to the question of describing all homomorphisms of the algebra $L^1(G)$ into the algebra $L^1(H)$ where G and H are two locally compact abelian groups. We shall treat this problem in a subsequent paper. As will be explained, the problem of Littlewood is closely connected to the problem of idempotent measures, and though the first is stated on the circle group the method of proof will be indispensable for the analysis on the more general class of abelian groups.

We now state the problem of Littlewood. Consider an exponential sum

$$\psi(x) = \sum_{j=1}^{N} e^{in_j x}.$$

Littlewood conjectured that

$$\int_0^{2\pi} |\psi(x)|\, dx > K \log N$$

for some constant K independent of N. No non-trivial lower bound for the L^1 norm of ψ has previously been obtained, and in particular it was unknown whether the L^1 norm necessarily tended to infinity as a function of N. Salem [6] has treated this problem and under suitable conditions on the growth of an infinite sequence of integers n_j he was able to show that for infinitely many N

$$\int_0^{2\pi} |\sum_{j=1}^{N} e^{in_j x}|\, dx > K(\log N)^{\frac{1}{2}}.$$

---

* Received January 9, 1959; revised version January 31, 1960.

This however tells us nothing about individual sums ψ. If n_j are taken equal to j, then ψ reduces essentially to the Dirichlet kernel and in that case the L^1 norm of ψ is of order of magnitude of $\log N$. Thus Littlewood's conjecture if true is certainly best possible. The case of the Dirichlet kernel may be thought of as the case of complete regularly for the n_j. The case of lacunary n_j, that is where $n_{j+1}/n_j > \lambda > 1$, may be thought of as the case of complete irregularity, and in that case it is well known that the exponentials e^{in_jx} behave similarly to independent functions and the L^1 norm is of the order of $N^{\frac{1}{2}}$. It thus would seem plausible that regular distribution of the integers n_j minimizes the L^1 norm and that Littlewood's conjecture is correct. In this direction we prove the following theorem:

THEOREM 1. *For some K, and all $N \geqq 3$, we have*

$$\int_0^{2\pi} \left| \sum_{j=1}^{N} c_j e^{in_jx} \right| dx > K(\log N/\log\log N)^{1/8} \tag{1}$$

whenever n_j are distinct integers and c_j satisfy $|c_j| \geqq 1$.

In particular we see that the right side of (1) tends to infinity as a function of N. We also shall prove (1) in the case of exponentials of several variables and still more generally, for sums of characters on arbitrary compact connected abelian groups.

To show the relation of Littlewood's problem to idempotent measures we make the following remarks. Suppose contrary to Theorem 1 that there exists a sequence of exponential sums ψ_n, ψ_n being a sum of n exponentials, such that $\| \psi_n \|_1$ remains bounded. By the weak compactness principle for measures it would then follow that the measures $\psi_n dx$ have a subsequence converging weakly to a measure μ, all of whose Fourier-Stieltjes coefficients are either zero or one. These so-called idempotent measures were analyzed in [3] and it was shown there that the sequence of coefficients which are equal to one must be essentially periodic. This, to some extent, tends to reinforce the belief that only periodic distribution of n_j can result in small L^1 norm for ψ.

In the first part of the paper we shall prove Theorem 1. The proof is elementary although at one point we appeal to some general principles in functional analysis. It is however completely free of considerations concerning general topological groups. In the second part, the method of proof of Theorem 1 will be used to analyze idempotent measures on a compact abelian group.

Part I.

1. We first consider the case in Theorem 1 when all the c_j are equal to one. We begin by reviewing some standard notation. If μ is a finite measure, we denote by $|\mu|$ the total variation of μ, that is

$$|\mu|(E) = \sup \sum_i |\mu(E_i)|$$

for all partitions of E into a disjoint union of sets E_i. If μ is absolutely continuous with respect to a non-negative measure dx, i. e.

$$\mu(E) = \int_E \alpha(x)\,dx$$

for all sets E, then we have

$$|\mu|(E) = \int_E |\alpha(x)|\,dx.$$

In general we have

$$\mu(E) = \int_E \theta(x)\,d\,|\mu|(x)$$

for all E, where $|\theta(x)| = 1$ a. e. with respect to $|\mu|$. We denote by $\|\mu\|$ the norm of μ, i. e. the total measure of $|\mu|$.

Let n_j, $1 \leqq j \leqq N$, be a sequence of increasing positive integers. Set

$$\mu = \sum_{j=1}^{N} e^{-in_jx}\,dx,$$

where dx is here used to denote $\frac{1}{2\pi}$ times ordinary Lebesgue measure on the interval $[0, 2\pi]$. Then we have

$$\begin{aligned}\int e^{inx}\,d\mu(x) &= 1 \text{ if } n = n_j \text{ for some } j\\ &= 0 \text{ otherwise.}\end{aligned}$$

Theorem 1 states that $\|\mu\|$ is greater than the right hand side of (1).

Before beginning the proof we make some heuristic remarks. First we remember that

$$\|\mu\| = \sup_{|\phi| \leqq 1} \left| \int \phi(x)\,d\mu(x) \right|. \tag{2}$$

By choosing $\phi = e^{in_1x}$, we obtain $\|\mu\| \geqq 1$. Actually unless $N = 1$, we must have $\|\mu\| > 1$. For if $\|\mu\| = 1$, then it is clear that $e^{in_1x}\mu(x)$ is a non-negative measure and hence since

$$\int e^{in_2x}\,d\mu(x) = \int e^{i(n_2-n_1)x} e^{in_1x}\,d\mu(x) = 1,$$

it follows that $e^{i(n_2-n_1)x}=1$ a. e. with respect to μ. Thus it follows that for any j, $n_j+n_2-n_1$ is also one of the original set of n_j. Since the set of n_j is finite, this is clearly absurd. This shows $\|\mu\|>1$. In the proof we shall use an analogous idea under the assumption that $\|\mu\|\leqq M$, to show that under certain conditions we will be led to a contradiction and that really $\|\mu\|>M+\delta$ for some quantity δ. This contradiction will be used to furnish functions ϕ to be used in (2). An inductive scheme will be set up to construct functions ϕ_k to be used in (2) each one showing that $\|\mu\|$ is greater than some number depending on k.

We now proceed to the proof of Theorem 1.

LEMMA 1. *Let μ be a measure on a space X such that $\|\mu\|\leqq M$, let N be an integer such that $N\geqq 2M^2$. Let f_j, $1\leqq j\leqq N$, be a set of functions such that*

a) $|f_j(x)|\leqq 1$ *for all* x,

b) $\int f_j(x)\,d\mu(x)=1$.

Then for some $i<j$ we have

$$\operatorname{Re}\int f_i\bar{f}_j\,d\,|\mu(x)|>1/2M.$$

Proof. We have $\mu(x)=\theta(x)\,|\mu|$, where $|\theta|\leqq 1$. Hence

$$N=\int\sum_{j=1}^{N}f_j(x)\theta(x)\,d\,|\mu(x)|.$$

By Schwarz's inequality

$$N^2\leqq\int|\sum_{j=1}^{N}f_j|^2\,d\,|\mu|\cdot\int|\theta|^2\,d\,|\mu|$$

or

$$N^2/M\leqq N\cdot M+2\sum_{i<j}\operatorname{Re}\int f_i\bar{f}_j\,d\,|\mu|.$$

Since $N\geqq 2M^2$,

$$1/2M\leqq(2/N^2)\sum_{i<j}\operatorname{Re}\int f_i\bar{f}_j\,d\,|\mu|.$$

Since the number of pairs $i<j$ is smaller than $N^2/2$, it follows that the conclusion of the lemma holds for at least one.

LEMMA 2. *Let μ be a measure on X, $\phi(x)$ and $g(x)$ functions such that $|\phi|\leqq 1$, $|g|\leqq 1$. Assume that*

a) $\int \phi \, d\mu(x) = M,\ M \geqq 1,$

b) $\left| \int g d \,|\, \mu(x) | \right| \geqq \alpha,\ 0 < \alpha < 1,$

c) $\int g\phi \, d\mu(x) = 0.$

Then

$$\| \mu \| \geqq M + \alpha^2/4M.$$

Proof. Since $d\mu(x) = \theta(x) d \,|\, \mu(x) |$, if we replace ϕ by $\phi\theta$ and μ by $|\mu|$, we see that μ may be assumed to be a non-negative measure. Let $\phi = \phi_1 + i\phi_2$, where ϕ_1 and ϕ_2 are real. Then

$$\int \phi_1 \, d\mu = M, \qquad \int \phi_2 \, d\mu = 0.$$

Set $\psi = \phi_1 + i \,|\, \phi_2 |$. Since $|\psi| \leqq 1$, we have

$$\| \mu \| \geqq \left| \int \psi \, d\mu \right| = \left| M + i \int | \phi_2 | \, d\mu \right|$$

or

$$\int | \phi_2 | \, d\mu \leqq (\| \mu \|^2 - M^2)^{\frac{1}{2}}.$$

Also,

$$\begin{aligned} \int | 1 - \phi | \, d\mu &\leqq \int (1 - \phi_1) \, d\mu + \int | \phi_2 | \, d\mu \\ &\leqq \| \mu \| - M + (\| \mu \|^2 - M^2)^{\frac{1}{2}}. \end{aligned}$$

But,

$$\alpha \leqq \left| \int (g - g\phi) \, d\mu \right| \leqq \int | 1 - \phi | \, d\mu.$$

Hence

$$\alpha \leqq \| \mu \| - M + (\| \mu \|^2 - M^2)^{\frac{1}{2}}. \tag{3}$$

If $\| \mu \| - M \geqq \alpha$, the lemma is clearly true. If not we may square both sides of (3) after transposing

$$(\alpha - \| \mu \| + M)^2 \leqq \| \mu \|^2 - M^2.$$

Hence

$$\begin{aligned} \| \mu \| &\geqq M + \alpha^2/2(\alpha + M) \\ &\geqq M + \alpha^2/4M. \end{aligned}$$

By combining Lemmas 1 and 2 we obtain:

LEMMA 3. *Given f_j, $1 \leqq j \leqq N$, and $\phi(x)$, where $|f_j(x)| \leqq 1$, $|\phi(x)| \leqq 1$, such that*

a) $\int f_j(x) \, d\mu(x) = 1,$

b) $\int \phi(x) \, d\mu(x) = M,\ M \geqq 1,$

c) *for all $i < j$,*

$$\int \phi(x) f_i(x) \bar{f}_j(x) \, d\mu(x) = 0,$$

then if $N \geqq 2(M+1)^2$, *we have*

$$\| \mu \| \geqq M + 1/64M^3.$$

For, if the conclusion were false then clearly $\| \mu \| \leqq M + 1$, so by Lemmas 1 and 2 we would have $\| \mu \| \geqq M + \alpha^2/4M$, where $\alpha = 1/2(M+1)$.

COROLLARY. *If* $\phi(x)$, $f_j(x)$ *are continuous functions satisfying the hypotheses of Lemma 3, then there exist constants* α, β_j, γ_{ij} *such that if*

$$\psi(x) = \alpha\phi(x) + \sum_j \beta_j f_j(x) + \sum_{i<j} \gamma_{ij} f_i(x) \bar{f}_j(x) \phi(x),$$

we have

$$|\psi(x)| \leqq 1, \text{ and } \int \psi(x)\, d\mu(x) = M + 1/64M^3.$$

Proof. Let V denote the linear subspace generated by $\phi(x)$, $f_j(x)$ and $\phi(x) f_i(x) \bar{f}_j(x)$. The measure μ induces a linear functional L on V. Let K be the norm of L on V. L can be extended to a functional on the space of all continuous functions with the same norm K. This new functional is given by a measure μ' satisfying the hypotheses of Lemma 3. Therefore by that lemma $K \geqq M + 1/64M^3$. This is precisely the conclusion of the corollary.

We need one more combinatorial lemma.

LEMMA 4. *Let* $E = \{n_1 > n_2 \cdots > n_N\}$ *be a collection of* N *integers,* r *a positive integer. If*

$$(2r^2)^{2s^2} \leqq N,$$

there exist sets P_k *and* T_k, $1 \leqq k \leqq s$, *of integers with the following properties:*

a) $T_k : \subset E, P_1 = \{n_1\}$,

b) *for all* $k \geqq 1$, $T_k = \{m_1^{(k)} > m_2^{(k)} > \cdots > m_r^{(k)}\}$ *and* $p + m_i^{(k)} - m_j^{(k)} \notin E$ *if* $p \in P_k$ *and* $i < j$,

c) P_{k+1} *is the union of* P_k, T_k *and all elements of the form* $p + m_i^{(k)} - m_j^{(k)}$, *where* $p \in P_k$, $i < j$.

Proof. We denote by $\alpha(k)$ the number of elements in P_k and by $\beta(k)$ the smallest integer such that $p \geqq n_{\beta(k)}$ for all p in P_k. Assume now that $P_1, \cdots, P_k$ and $T_1, \cdots, T_{k-1}$ have been chosen and satisfy a), b), and c). We shall now define the set T_k. Set $m_1^{(k)} = n_1$. Assume that $m_1^{(k)}, \cdots, m_t^{(k)}$ have been chosen and satisfy b), where $m_i^{(k)} = n_{\delta(i)}$ for $i \leqq t$. We then define $m_{t+1}^{(k)} = n_{\delta(t+1)}$, where $\delta(t+1)$ is the smallest number such that this choice of $m_{t+1}^{(k)}$ does not violate condition b). For each p in P_k and $i \leqq t$, the number of choices of $m_{t+1}^{(k)} < m_t^{(k)}$ such that

$$p + m_i^{(k)} - m_{t+1}^{(k)} \in P_k$$

is clearly less than $\beta(k)$. Thus it follows that if $m_{t+1}{}^{(k)}$ avoids at most $t\cdot\beta(k)\cdot\alpha(k)$ possible choices, b) will be satisfied. That is

$$\delta(t+1)-\delta(t)\leqq 1+t\cdot\alpha(k)\cdot\beta(k).$$

It thus follows that

$$\delta(r)\leqq r+\tfrac{1}{2}r(r-1)\alpha(k)\beta(k).$$

Having thus defined T_k, we define P_{k+1} by means of c). Clearly we have

$$\beta(k+1)\leqq 2r^2\alpha(k)\beta(k).$$

The number of elements in P_{k+1} is smaller than

$$r+\alpha(k)+\tfrac{1}{2}r(r-1)\alpha(k)$$

or

$$\alpha(k+1)\leqq 2r^2\alpha(k).$$

We thus conclude that

$$\alpha(k)\leqq(2r^2)^k$$

and

$$\begin{aligned}\beta(k)&\leqq(2r^2)^{2+3+\cdots+(k+1)}\\&\leqq(2r^2)^{k\cdot(k+3)/2}\\&\leqq(2r^2)^{2k^2}.\end{aligned}$$

Clearly the sets P_s, T_s can be constructed if

$$\beta(s)\leqq N$$

or

$$(2r^2)^{2s^2}\leqq N.$$

We now return to the case $\mu=\sum_{j=1}^{N}e^{-in_jx}\,dx$. Assume that $\|\mu\|\leqq M$, $M\geqq 1$, and set $r=[2M^2]+1$. We shall define functions ϕ_k, $|\phi_k|\leqq 1$, $1\leqq k\leqq s$, such that each ϕ_k is a linear combination of exponentials of the form e^{inx}, where n belongs to P_k. Furthermore,

$$\int\phi_k\,d\mu(x)=1+(k-1)/64M^3. \tag{4}$$

Set $\phi_1=e^{in_1x}$. If ϕ_k has been defined, then by the corollary to Lemma 3, if we set $f_j(x)=\exp(im_j{}^{(k)}x)$, we clearly have that there exists a function ψ, which we now denote by ϕ_{k+1}, satisfying

$$\int\phi_{k+1}\,d\mu(x)=1+(k-1)/64M^3+1/64M^3,\qquad|\phi_{k+1}|\leqq 1,$$

since we must always have

$$\int \phi_k \, d\mu(x) \leqq M.$$

Furthermore it is clear that ϕ_{k+1} is a linear combination of e^{inx}, where n lies in P_{k+1}. It thus follows that

$$1 + (s-1)/64M^3 \leqq M, \tag{5}$$

where s is any number that satisfies

$$(2(2M^2+1)^2)^{2s^2} \leqq N \tag{6}$$

or

$$2s^2 \log[2(2M^2+1)^2] \leqq \log N. \tag{7}$$

Since (5) is false for $s = C_1 M^4$, we have that

$$C_2 M^8 \log(2M^2+1) \geqq \log N$$

or

$$C_3 M^8 \log M \geqq \log N$$

for $N \geqq 3$. Thus if we define M_1 as the solution of the equation

$$C_3 M_1{}^8 \log M_1 = \log N,$$

clearly

$$M_1 \leqq C_4 (\log N)^{1/8}$$

so that

$$M_1{}^8 \geqq C_5 \log N / \log \log N$$

and hence $M \geqq K(\log N/\log\log N)^{1/8}$ which is precisely Theorem 1 in the case of coefficients equal to one.

2. We now consider the general case of Theorem 1.

LEMMA 1′. *Let f_j and μ be as in Lemma 1. Assume that*

$$\int f_j(x)\, d\mu(x) = c_j, \qquad |c_j| \geqq 1.$$

Then we still have for some $i < j$

$$\left| \int f_i \bar{f}_j \, d|\mu| \right| > 1/2M.$$

Proof. Set $f_j' = (|c_j|/c_j) f_j$. Then

$$\int f_j' \, d\mu = |c_j|.$$

As in the proof of Lemma 1, we have

$$N \leqq \int \sum_{j=1}^{N} f_j'(x)\theta(x)\, d|\mu|.$$

Exactly as before we deduce that for some $i < j$

$$\operatorname{Re} \int f_i' \bar{f}_j' \, d|\mu| > 1/2M.$$

This clearly implies the lemma.

From this point on the proof of Theorem 1 proceeds precisely as in the case already discussed.

It might be pointed out here that in the special case of the Dirichlet kernel our method can be improved by a sharper version of Lemma 4 to yield Theorem 1 with the exponent 1/4 replacing 1/8. However in this case one has the more powerful theorem of Hardy-Littlewood, namely [7],

$$\sum_{n=0}^{\infty} |c_n|/(n+1) \leqq K \int_0^{2\pi} |\sum_{j=0}^{\infty} c_j e^{in_j x}| \, dx.$$

THEOREM 2. *For any compact connected abelian group G we have*

$$\int |\sum_{j=1}^{N} c_j f_j| \, dm > K(\log N/\log\log N)^{1/8}, \qquad |c_j| \geqq 1, N \geqq 3,$$

where f_j are distinct characters on G and K is an absolute constant independent of G.

In particular Theorem 1 holds for exponentials in more than one variable. The condition that G be connected merely implies that $\hat{G}$, the character group of G, has no elements of finite order. To prove Theorem 2 we first need a definition.

Definition. Let H and H' be two additive discrete abelian groups, x_i a finite set in H, ϕ a one-one mapping of x_i into H'. ϕ is said to be an n-isomorphism if for all integers n_i, $|n_i| \leqq n$, we have $\sum n_i x_i = 0$ if and only if $\sum n_i \phi(x_i) = 0$.

LEMMA 5. *If $f_1, \cdots, f_N$ are characters on G, n a positive integer, there exist exponentials $g_1, \cdots, g_N$ on the circle such that the map ϕ, defined by $\phi(f_i) = g_i$ is an n-isomorphism from f_i into the character group of the circle, namely, the integers.*

Proof. Let H be the group generated by $f_1, \cdots, f_N$. By the fundamental theorem for finitely generated abelian groups it follows that we have

$$f_i = \sum_{j=1}^{p} d_{ij} f_j',$$

where the f'_j are linearly independent elements of $\hat{G}$, and d_{ij} are integers.

(Here we consider $\hat{G}$ as an additive group). Notice that $p < N$. Let g_j' be a set of integers so that if

$$\sum_{j=1}^{p} z_j g_j' = 0, \text{ then } |z| \geqq u$$

for some j, where u is a number to be chosen in a moment. Such a choice is always possible. Set

$$\phi(f_i) = \sum_{j=1}^{p} d_{ij} g_j'.$$

If $\sum c_i f_i = 0$, then clearly $\sum_i c_i d_{ij} = 0$ for each j. Thus $\sum c_i \phi(f_i) = \sum_j (\sum_i c_i d_{ij}) g_j' = 0$. Conversely if $\sum c_i \phi(f_i) = 0$, and $|c_i| \leqq n$, we have $\sum_j (\sum_i c_i d_{ij}) g_j' = 0$. If u is now chosen greater than $n\ N \cdot D$, where $D = \operatorname{Max} |d_{ij}|$, it follows then that $\sum_i c_i d_{ij} = 0$ and so $\sum c_i f_i = 0$.

Lemma 4 now tells us that given any set E of N integers we can find subsets T_k of E having certain properties. All these properties are of the form that certain linear combinations of elements in E do not lie in E, or equivalently that certain linear combinations are not zero. There is an upper bound to the magnitude of the coefficients in these linear combinations, say z. If now g_j are ordinary exponentials such that the map defined by $\phi(f_j) = g_j$ is a z-isomorphism, it is clear that the proof of Theorem 1 for the case of the g_j may be carried over word for word to the case of the f_j. Thus Theorem 2 is proved.

If we apply Theorem 2 to the case of the Bohr compactification of the real line we obtain that

$$\lim_{L\to\infty} (1/2L) \int_{-L}^{L} \left| \sum_{j=1}^{N} e^{i\lambda_j x} \right| dx > K(\log N/\log\log N)^{1/8},$$

where λ_j are arbitrary distinct real numbers.

Part II.

1. Let G be a locally compact abelian group, $\hat{G}$ its dual group. We denote by $M(G)$ the algebra of finite measures on G under convolution, that is

$$(\mu * \nu)(E) = \int\int_D d\mu(x)\, d\nu(y),$$

where D is the set of all (x, y) where $xy \in E$. $M(G)$ is a Banach algebra with the usual norm $\|\mu\|$. The set of absolutely continuous measures forms

a subalgebra which may be identified with $L^1(G)$. The Fourier transform of μ is defined on $\hat{G}$

$$\hat{\mu}(\chi)=\int\chi(x)\,d\mu(x).$$

We have the fundamental fact that the Fourier transform of $\mu*\nu$ is the product of the Fourier transforms of μ and ν.

In this part we shall investigate measures μ which satisfy $\mu*\mu=\mu$. This problem has been treated by several authors. It has been solved by Helson in the case of the circle group, and by Rudin in the case of the n-dimensional torus [3], [5]. Both these authors use either real or complex variables in such a manner as to seemingly prevent their application to arbitrary groups. The approach we use is similar to the method used in Part I.

In the case in which G is the Bohr compactification of the real line, the problem in a different but equivalent formulation was treated by Helgason [2], who obtained certain necessary conditions.

Let μ now be an idempotent measure. Since $\hat{\mu}^2=\hat{\mu}$, it follows that $\hat{\mu}$ only assumes the values zero and one. It is well known that $\hat{\mu}$ is uniformly continuous [4], and hence $\hat{\mu}$ is invariant under translation by some open set U containing the identity. Thus the subgroup of $\hat{G}$ which leaves $\hat{\mu}$ invariant under translation is an open subgroup. Now if χ is a character, the Fourier transform of $\chi\mu$ is precisely $\hat{\mu}$ translated by χ. Since $\hat{\mu}$ determines μ [4], it follows that μ is invariant under multiplication by all χ in some open subgroup S of G. Thus the support of μ lies in the annihilator of S, which by the duality theorem must be compact. Thus we have shown that the support of an idempotent measure is always contained in a compact subgroup. This result was obtained by Rudin in [5]. Thus we may restrict our attention to compact G. In this case of course $\hat{G}$ is discrete. If H is a subgroup of G, we have the canonical may $\phi\colon G\to G/H$. ϕ induces a map ϕ_* of $M(G)$ into $M(G/H)$, namely $(\phi_*\mu)(E)=\mu(\phi^{-1}(E))$. If K is the subgroup of $\hat{G}$ annihilating H, then K is the dual group to G/H. If χ belongs to K, it is easy to verify that

$$(\widehat{\phi_*\mu})(\chi)=\int\chi\,d\mu.$$

Thus the Fourier transform of $\phi_*\mu$ is the restriction of $\hat{\mu}$ to K. Hence ϕ_* is a homomorphism from $M(G)$ into $M(G/H)$.

If μ is an absolutely continuous idempotent, since $\hat{\mu}$ must then vanish at infinity, it follows that $\hat{\mu}$ is one only finitely often and hence

$$\mu=(\chi_1+\cdots+\chi_n)\,dm,$$

where χ_j are characters on G and dm is Haar measure. Thus the analysis of absolutely continuous idempotents is trivial. If H is a closed subgroup of G, clearly $M(H)$ is a subalgebra of $M(G)$. An absolutely continuous idempotent on H can be regarded as a measure on G and is still clearly an idempotent. Such an idempotent we call a primitive idempotent. For μ an idempotent, $\hat{\mu}$ is the characteristic function of a set E. If $\hat{\mu}_1$ and $\hat{\mu}_2$ are the characteristic functions of E_1 and E_2, then the Fourier transforms of $\mu_1 * \mu_2$ and $\delta - \mu_1$, where δ is the point mass of one concentrated at the identity, are the characteristic functions of $E_1 \cap E_2$ and the complement of E_1 respectvely. Similarly $\mu_1 + \mu_2 - \mu_1 * \mu_2$ has as its Fourier transform the characteristic function $E_1 \cup E_2$. These three operations we shall call the intersection, complementation and union of idempotent measures.

If μ is Haar measure of a subgroup H of G, then $\hat{\mu}$ is the characteristic function of K, the subgroup of $\hat{G}$ which annihilates H. Since the Fourier transform of $\chi\mu$ is $\hat{\mu}$ translated by χ, it follows that for μ a primitive idempotent, $\hat{\mu}$ is the characteristic function of a set E which lies in the Boolean ring generated by cosets of subgroups of $\hat{G}$. Conversely, it is clear that if E is a set which lies in the Boolean ring generated by cosets of subgroups of $\hat{G}$, then the characteristic function of E is the Fourier transform of an idempotent measure.

THEOREM 3. *Every idempotent measure on a compact abelian group G lies in the Boolean ring generated by primitive idempotens. Equivalently, if μ is an idempotent measure, $\hat{\mu}$ is the characteristic function of a set E which lies in the Boolean ring generated by cosets of subgroups of $\hat{G}$.*

The second part of Theorem 3 is easily seen to be equivalent to the first by the above observations.

It is sufficient to prove

THEOREM 3′. *If μ is an idempotent measure, then μ is a linear combination of primitive idempotents.*

For, if $\mu = \sum \alpha_j \mu_j$, where μ_j are primitive, it follows that $\hat{\mu} = \sum \alpha_j \hat{\mu}_j$. Now if $\hat{\mu}_j$ is the characteristic function of E_j, it is clear that $\hat{\mu}$ is the characteristic function of a set E which lies in the Boolean ring generated by E_j. Thus E lies in the Boolean ring generated by cosets of subgroups.

We now state the main lemma.

MAIN LEMMA. *If μ is an idempotent measure, either μ is absolutely continuous or else there exists a subgroup H of G which is of infinite index in G and such that $|\mu|(H) \neq 0$.*

To show how the Main Lemma implies the theorem we introduce some preliminary notions. Let H be a closed subgroup of G. Denote by μ_H that part of the measure μ which lies on any coset of H. More precisely,

$$\mu_H(E) = \sum_{gH} \mu(E \cap gH), \tag{1}$$

where gH ranges over all cosets of H. Of course the sum in (1) is at most a countable one. It is easy to verify

$$\| \mu \| = \| \mu_H \| + \| \mu - \mu_H \|. \tag{2}$$

If $\mu_H = 0$, then for any ν and any E contained in a coset of H we have

$$(\mu * \nu)(E) = \int \mu(t^{-1}E)\, d\nu(t) = 0.$$

Hence $(\mu * \nu)_H = 0$. Thus the set of μ such that $\mu_H = 0$ is an ideal. If $\mu_H = \mu$ and $\nu_H = \nu$, we have

$$\mu = \sum_j \mu_j, \qquad \nu = \sum \nu_j,$$

where μ_j and ν_j are measures with their supports in cosets of H. Hence

$$\mu * \nu = \sum_{i,j} \mu_i * \nu_j.$$

Thus it follows that $(\mu * \nu)_H = \mu * \nu$, since $\mu_i * \nu_j$ also has its support contained in a coset of H. More generally, if

$$\mu = \mu_H + \mu', \qquad \nu = \nu_H + \nu',$$

we have

$$\mu * \nu = \mu_H * \nu_H + \mu' * (\nu_H + \nu') + \nu' * \mu_H. \tag{3}$$

By the above remarks it follows from (3) that $(\mu * \nu)_H = \mu_H * \nu_H$. Thus the map $\mu \to \mu_H$ is a homomorphism.

The mapping $\mu \to \mu_H$, where H is the identity element alone, yields the discrete part of μ, and was used by Helson in [3]. Rudin in [5] also uses this map for general H.

If μ is an idempotent, clearly so is μ_H. We now put a new topology on G as follows. A set is open whenever its intersection with every coset of H is open. G is clearly still locally compact under the new topology. The measure μ_H can now be regarded as a measure under the new topology. Now it follows that the support of μ_H is contained in a compact subgroup H'. The group $H_0 = H \cdot H'$ is still a compact subgroup so that it may intersect only a finite number of cosets of H. Thus $[H_0 : H]$, the index of H in H_0, is finite, and we have shown that the support of μ_H is contained in a subgroup H_0 in which H is of finite index.

Assume now that the Main Lemma is true. If μ is a measure such that $\hat{\mu}$ takes only a finite number of values, the conclusion of the lemma is still true. For, if $\hat{\mu}$ takes only the non-zero values $\alpha_1, \cdots, \alpha_k$, we have $\mu = \sum \alpha_k \mu_k$, where $\mu_k = P_k(\mu)$ and $P_k(x)$ is a polynomial such that $P_k(\alpha_k) = 1$, $P_k(0) = 0$, and $P_k(\alpha_l) = 0$ if $k \neq l$ and μ_k are idempotents. Furthermore $\mu_k * \mu_l = 0$ if $k \neq l$. Now if μ is not absolutely continuous we must have, say, μ_1 not absolutely continuous. Thus for some H of infinite index $\mu_{1H} \neq 0$. Also $\mu_H = \sum \alpha_k \mu_{kH}$ and $\mu_{kH} * \mu_{lH} = 0$ if $k \neq l$. Thus $\mu_H * \mu_{1H} = \alpha_1 \mu_{1H} \neq 0$ so that μ_H is not zero. By the above argument μ_{kH} has support contained in a subgroup $H_0^{(k)}$, where $[H_0^{(k)} : H]$ is finite. Thus μ_H has its support contained in $\prod_k H_0^{(k)} = \tilde{H}$, where $[\tilde{H} : H]$ is finite. Thus μ_H has its support contained in a subgroup $\tilde{H}$ of infinite index and $|\mu|$ does not vanish on $\tilde{H}$, since $\mu_H \neq 0$.

Assume that we know that if $\| \mu \| \leqq n$ and $\hat{\mu}$ takes only integral values, that μ is a linear combination of primitive idempotents. Clearly this is true for $n = 0$. Observe that if $\hat{\mu}$ takes no values other than $w_1, \cdots, w_k$, then

$$(\mu - w_1) * \cdots * (\mu - w_k) = 0.$$

Hence

$$(\mu_H - w_1) * \cdots * (\mu_H - w_k) = 0$$

so that $\hat{\mu}_H$ takes at most these values. Let μ be a measure such that $\hat{\mu}$ takes only integer values and $\| \mu \| \leqq n + 1$, μ not absolutely continuous. We may also assume that the support of μ is not contained in any proper subgroup of G since otherwise we need only replace G by the closed group generated by the support of μ. By the lemma and the above remarks, for some H of infinite index in G, $\mu_H \neq 0$. Since $\hat{\mu}_H$ takes only integral values, $\| \mu_H \| \geqq 1$. By (2)

$$\| \mu_H \| \leqq n, \quad \text{and} \quad \| \mu - \mu_H \| \leqq n.$$

since $\mu \neq \mu_H$. By hypothesis both these measures are linear combinations of primitive idempotents and so Theorem 3′ and hence Theorem 3 is proved.

2. In this section we shall prepare the proof of the Main Theorem. From now on we assume that μ is an idempotent measure which is not absolutely continuous, and that $\hat{\mu}$ is the characteristic function of E, a subset of $\hat{G}$.

Definition. The finite set S is said to be a set of periods or a P-set for E if whenever f belongs to E there is some g in S such that fg belongs to E, except for at most finitely many f.

THEOREM 4. *There exists a P-set for E disjoint from any given finite set A.*

Proof. The proof of Theorem 4 follows from a close examination of the proof of Theorem 1. Assume that $\| \mu \| \leqq M$. It then follows, as in the proof of Theorem 1, that there cannot exist sets T_k, P_k satisfying the properties of Lemma 4, $1 \leqq k \leqq s$, if s equals CM^4. Assume Theorem 4 is false. We shall now define sets T_k satisfying Lemma 4 which will be a contradiction. Define T_s to be any collection of elements of E, $m_1^{(s)}, \cdots, m_r^{(s)}$ such that $m_i^{(s)} - m_j^{(s)}$ for $i < j$ is never a member of the finite set A. This is clearly possible. Call the set of elements $m_i^{(s)} - m_j^{(s)}$, $i < j$, B_1. Since B_1 is not a P-set there must be an infinite number of elements n in E such that $n + b \notin E$ for any $b \in B_1$. Thus there are r elements $m_1^{(s-1)}, \cdots, m_r^{(s-1)}$ such that $m_i^{(s-1)} + b \notin E$ for b in B_1 and furthermore if we set B_2 equal to the union of B_1, $m_i^{(s-1)} - m_j^{(s-1)}$ for $i < j$, and $b + m_i^{(s-1)} - m_j^{(s-1)}$ for b in B_1 and $i < j$, then B_2 is disjoint from A. This is possible because there are an infinite number of choices for $m_i^{(s-1)}$ and A is only finite. Set

$$T_{s-1} = \{m_1^{(s-1)}, \cdots, m_r^{(s-1)}\}.$$

Having defined $T_s, T_{s-1}, \cdots, T_{s-k}$, as well as $B_1, B_2, \cdots, B_{k+1}$ we define T_{s-k-1} and B_{k+2} as follows. Since B_{k+1} is not a P-set there are as before infinitely many n in E such that $n + b \notin E$ for b in B_{k+1}. Choose

$$T_{s-k-1} = \{m_1^{(s-k-1)}, \cdots, m_r^{(s-k-1)}\}$$

such that first, $n + b \notin E$ for n in T_{s-k-1} and b in B_{k+1} and such that if B_{k+2} is defined as the union of B_{k+1}, $m_i^{(s-k-1)} - m_j^{(s-k-1)}$ where $i < j$, and $b + m_i^{(s-k-1)} - m_j^{(s-k-1)}$ for b in B_{k+1} and $i < j$, then B_{k+2} is disjoint from A. We now maintain that with this choice of T_k and P_1 chosen to be the set consisting only of one element n in E, where $n + b \notin E$ for b in B_s, that the resulting T_k and P_k satisfy the hypotheses of Lemma 4. To show this we must prove that $P_k + \{m_i^{(k)} - m_j^{(k)}\}$, $i < j$, has no elements in E. This is certainly true if $P_k + B_{s-k+1}$ has no element in E. By the definition of P_k this will be true if neither $P_{k-1} + B_{s-k+1}$, nor $T_{k-1} + B_{s-k+1}$, nor

$$P_{k-1} + B_{s-k+1} + \{m_i^{(k-1)} - m_j^{(k-1)}\}$$

have elements in E. The second set clearly has no element in E by definition of T_{k-1}. Thus we need only show that

$$P_{k-1} + \{B_{s-k+1} \cup \{B_{s-k+1} + m_i^{(k-1)} - m_j^{(k-1)}\}\}$$

have no element in E. This will be true if $P_{k-1} + B_{s-k+2}$ has no element in E by definition of B_{s-k+2}. Thus we finally see by induction that $P_k + m_i^{(k)} - m_j^{(k)}$ has no element in E provided $P_1 + B_s$ has no element in E, which is true.

Thus this choice of T_k and P_k satisfy Lemma 4 and the proof of Theorem 1 would now show that $\| \mu \| > M$ which is a contradiction and Theorem 4 is proved.

COROLLARY. *In the case where G is the circle group, the set S of periods may be chosen to be positive integers.*

This is so because if we order each set T_k in descending order, the numbers $m_i^{(k)} - m_j^{(k)}$ for $i < j$ are all positive and hence so are all the sets B_k.

THEOREM 5 (Helson). *The Main Lemma holds if G is the circle group.*

To prove this we need the following rather simple theorem of Wiener [7]. If μ is a measure on the circle with no point mass, then if $c_n = \int_0^{2\pi} e^{inx}\, d\mu$, we have

$$\lim_{N\to\infty} (1/N) \sum_{n=-N}^{N} | c_n | = 0.$$

If μ is an idempotent neither absolutely continuous nor having point mass, then the set E is an infinite set of integers which by Wiener's theorem has gaps of arbitrarily large length. That is, there exist n_j in E such that $n_j + k$ is not in E for $1 \leqq k \leqq j$. Clearly E cannot then have a P-set of positive integers and so the theorem is proved.

LEMMA 6. *Let μ be an arbitrary measure, ϕ, g, and h functions such that $|\phi| \leqq 1$, $|g| \leqq 1$, $|h| \leqq 1$. If*

$$\int \phi\, d\mu = M, \qquad M \geqq 1,$$
$$\int \phi g\, d\mu = M,$$
$$\int h\, d\mu = 1, \qquad \int gh\, d\mu = 0,$$

then

$$\| \mu \| \geqq M + \delta(M),$$

where $\delta(M)$ is a positive decreasing function of M.

Proof. If $\mu = \theta(x) |\mu|$, where $|\theta| = 1$, we see that replacing ϕ by $\phi\theta$, h by $h\theta$, and μ by $|\mu|$ shows that we may assume μ is a non-negative measure. If $\phi = \phi_1 + i\phi_2$, we have

$$\int \phi_1\, d\mu = M, \qquad \int \phi_2\, d\mu = 0.$$

Thus

$$\| \mu \| \geqq | \int (\phi_1 + i\, |\phi_2|)\, d\mu | = | M + i \int |\phi_2|\, d\mu |$$

or

$$\int |\phi_2|\, d\mu \leqq (\| \mu \|^2 - M^2)^{\frac{1}{2}}.$$

Hence

$$
(4) \qquad \begin{aligned} \int |1-\phi|\, d\mu &\leqq \int (1-\phi_1)\, d\mu + \int |\phi_2|\, d\mu \\ &\leqq \|\mu\| - M + (\|\mu\|^2 - M^2)^{\frac{1}{2}}. \end{aligned}
$$

We rewrite (4) more simply as

$$
(5) \qquad \int |1-\phi|\, d\mu \leqq R_1(\|\mu\| - M),
$$

where $R_1(t) \to 0$ as $t \to 0$. (Here we regard M as fixed.) Also,

$$
\left| \int g\, d\mu - \int \phi g\, d\mu \right| \leqq \int |1-\phi|\, d\mu \leqq R_1(\|\mu\| - M).
$$

Thus,

$$
\int g_1\, d\mu \geqq M - R_1(\|\mu\| - M),
$$

where $g = g_1 + ig_2$. Now,

$$
\begin{aligned} \|\mu\| &\geqq \left| \int (g_1 + i|g_2|)\, d\mu \right| \\ &\geqq ([M - R_1(\|\mu\| - M)]^2 + [\int |g_2|\, d\mu]^2)^{\frac{1}{2}} \end{aligned}
$$

so that it follows easily that

$$
(6) \qquad \int |g_2|\, d\mu \leqq R_2(\|\mu\| - M),
$$

where $R_2(t) \to 0$ as $t \to 0$. Thus

$$
(7) \qquad \begin{aligned} \int |1-g|\, d\mu &\leqq \int (1-g_1)\, d\mu + \int |g_2|\, d\mu \\ &\leqq M - \int g_1\, d\mu + R_2(\|\mu\| - M) \\ &\leqq R_1(\|\mu\| - M) + R_2(\|\mu\| - M) \\ &\leqq R_3(\|\mu\| - M), \end{aligned}
$$

where $R_3(t) \to 0$ as $t \to 0$. But,

$$
1 = \int (h - gh)\, d\mu \leqq \int |1-g|\, d\mu \leqq R_3(\|\mu\| - M).
$$

Thus

$$
\|\mu\| \geqq M + \delta(M),
$$

where $\delta(M)$ is positive and is easily seen to depend continuously on M. Thus it may certainly be taken to be decreasing.

COROLLARY. *Given $M \geqq 1$, there exist constants $\alpha_1(M), \cdots, \alpha_4(M)$ such that if μ, ϕ, g, h satisfy the hypotheses of Lemma 6, then if*

$$
\psi = \alpha_1\phi + \alpha_2\phi g + \alpha_3 h + \alpha_4 gh,
$$

we have $|\psi| \leqq 1$, and $\int \psi\, d\mu = M + \delta(M)$.

2

Proof. Let X denote the product of three copies of the unit disc $|z| \leqq 1$. Let p_i be the function on X defined by $p_i(z_1, z_2, z_3) = z_i$. Let V be the subspace of $C(X)$ spanned by p_1, p_1p_2, p_3, and p_2p_3. Let L be a linear functional on V defined by $L(p_1) = M$, $L(p_1p_2) = M$, $L(p_3) = 1$, $L(p_2p_3) = 0$. Let K be the norm of L on V. L can be extended to a functional on all of $C(X)$ with norm K and is therefore given by a measure μ on X, where $\| \mu \| = K$. By Lemma 6, if we substitute p_1, p_2, and p_3 for ϕ, g, and h respectively, it follows that $\| \mu \| = K \geqq M + \delta(M)$. Thus there are constants α_1, α_2, α_3 and α_4 such that

$$(8) \qquad | \alpha_1 p_1 + \alpha_2 p_1 p_2 + \alpha_3 p_3 + \alpha_4 p_2 p_3 | \leqq 1$$

and

$$(9) \qquad \alpha_1 M + \alpha_2 M + \alpha_3 = M + \delta(M).$$

By the definition of p_i, (8) holds if p_i are arbitrary numbers less than or equal to one in absolute value. Thus the function ψ in the corollary satisfies $|\psi| \leqq 1$ and (9) clearly implies $\int \psi \, d\mu = M + \delta(M)$.

LEMMA 7. *Let $K_n \subset K_{n+1}$ be a strictly increasing sequence of finite subgroups of G, μ an idempotent measure on G, and $\hat{\mu}$ the characteristic function of E. If for a sequence g_n it is true that E contains g_nK_n, then $|\mu|$ does not vanish on some subgroup H of infinite index in G.*

Proof. Let H_n be the subgroup of G which annihilates K_n. Then $H_n \supset H_{n+1}$ and the measure $g_n\mu$ is an idempotent whose Fourier transform is the characteristic function of a set E_n, where $E_n \supset K_n$. The map ϕ_n from $G \to G/H_n$ induces ϕ_{n^*} from $M(G)$ into $M(G/H_n)$. As was shown earlier the Fourier transform of $\phi_{n^*}(g_n\mu)$ is the Fourier transform of $g_n\mu$ restricted to K_n, or in other words, identically one. Thus $\phi_{n^*}(g_n\mu)$ is the point measure of mass one at the origin, which implies that $g_n\mu$ satisfies

$$\int_{H_n} g_n \, d\mu = 1.$$

Thus $|\mu|(H_n) \geqq 1$. If $H = \bigcap_n H_n$, $|\mu|(H) \geqq 1$ and H is of infinite index, which proves the lemma.

LEMMA 8. *Let μ and E be as before. Then either μ has mass on a subgroup of infinite index or for each f in G, E contains only finitely many elements of the form f^j.*

Proof. We may clearly assume f is of infinite order. Let H be the subgroup of G annihilating f. By duality we know that G/H is the circle

group. As before ϕ denotes the map from G to G/H and ϕ_* the induced map from $M(G)$ into $M(G/H)$. The Fourier transform of $\phi_*(\mu)$ is one infinitely often by the hypothesis and so by Theorem 5, $\phi_*(\mu)$ has mass on a finite subgroup D of the circle. This implies that μ has mass on $\phi^{-1}(D)$ which is of infinite index in G.

3. After these preparations we now begin the proof of the Main Lemma. To illustrate the method we shall first treat the case where $\hat{G}$ has no elements of finite order. We assume $\|\mu\| \leqq M$, and μ has no mass on a subgroup of infinite index.

LEMMA 9. *For any f in $\hat{G}$ of infinite order and arbitrary g, the number of elements of the form gf^j which lie in E is bounded by a number N depending only on M.*

Proof. As in the proof of Lemma 8, it follows that $\phi_*(g\mu)$ is a finite exponential sum. Since $\|\phi_*(g\mu)\| \leqq M$, the number of terms is bounded by some number $N(M)$ by Theorem 1.

We shall now define functions $\phi_j^{(k)}$, $1 \leqq j \leqq \alpha(k)$, $|\phi_j^{(k)}| \leqq 1$ such that for all f in E and all k there exists a j so that $\int f\phi_j^{(k)} d\mu = 1 + k\delta(M)$. Since this is impossible for $k \geqq M/\delta(M)$ we will have obtained a contradiction. We may clearly take $\alpha(0) = 1$ and $\phi_1^{(0)} = 1$. Assume $\phi_j^{(k)}$ have been defined. Let S_m be a collection of disjoint P-sets, where $1 \leqq m \leqq \alpha(k) + 1$. This is possible because of Theorem 4. In the definition of a P-set we allowed for a finite number of exceptional elements. However by adjoining a finite number of elements disjoint from any given finite set it is clear that we may assume that there are no exceptional values. Thus, if $f \in E$ we have for each m, an element g_m such that $fg_m \in E$, and no two elements g_m are equal. Thus there are functions $\phi_{j(m)}^{(k)}$ such that $\int fg_m\phi_{j(m)}^{(k)} d\mu = 1 + k\delta(M)$. Since m ranges through $\alpha(k) + 1$ values it follows that for two values m_1 and m_2, $j(m_1) = j(m_2)$. Set $g = g_{m_1}^{-1}g_{m_2}$. Since g is not the identity element, it is of infinite order. By Lemma 9, for some $0 \leqq n \leqq N$ we have $fg^n \in E$, but $fg^{n+1} \notin E$. We now apply Lemma 6, where the roles of ϕ, g, and h are played by $fg_{m_1}\phi_{j(m_1)}^{(k)}$, g, and fg^n respectively. The role of M is now played by $1 + k\delta(M)$. Since we must have $1 + k\delta(M) \leqq M$ and $\delta(M)$ is a decreasing function we can clearly multiply the function ψ of the corollary by $\{1 + k\delta(M) + \delta(M)\}/\{1 + k\delta(M) + \delta(1 + k\delta(M))\}$ and still have $|\psi| \leqq 1$. Thus it follows by the corollary that there are absolute constants α_1, α_2, α_3 and α_4 so that if

$$\psi = f(\alpha_1 g_{m_1}\phi_{j(m_1)}^{(k)} + \alpha_2 g_{m_2}\phi_{j(m_1)}^{(k)} + \alpha_3 g^n + \alpha_4 g^{n+1}),$$

$|\psi| \leqq 1$, $\int \psi \, d\mu = 1 + (k+1)\delta(M)$. Now it is clear that only finitely many functions ψ/f can arise in this manner. Thus if we define $\phi_j^{(k+1)}$ to run through all these functions we see that the construction is completed. Thus the Main lemma is proved in the case where $\hat{G}$ has no elements of finite order.

In the case G has elements of finite order we will use Lemma 7 as well as the ideas already employed in the above proof. Set $s = [M/\delta(M)] + 1$. For each $k \geqq 1$ and $0 \leqq p \leqq s$ we will construct finite collections U_k of cosets of finite subgroups of $\hat{G}$ and finite collections of functions Φ_p^k. These will satisfy the following properties.

A. For each f in E and $k \geqq 1$, there is a coset aK in U_k such that faK is contained in E. Conversely, if aK belongs to U_k, there is an f in E such that faK is contained in E. Also, $U_k \subset U_{k+1}$.

B. For each f in E and $k \geqq 1$, there exists p, $0 \leqq p \leqq s$, and a function ϕ in Φ_p^k such that $|\phi| \leqq 1$, and $\int f\phi \, d\mu = 1 + p\delta(M)$. Also $\Phi_p^k \subset \Phi_p^{k+1}$.

C. The subgroup K of condition A and the integer p of condition B may be chosen so as to satisfy the following. There are cosets $a_0K, a_1K_1, \cdots, a_{k-p}K_{k-p}$, where a_jK_j belongs to U_{k-j}, and $K \supset K_1 \supset \cdots \supset K_{k-p}$, where the inclusions are proper.

We define U_1 to consist of the one element group alone, while Φ_0^1 consists only of the function 1. For $p \geqq 1$, Φ_p^1 is empty. Assume now that U_k and Φ_p^k have been defined. We introduce the notation $\operatorname{ord}(T)$ to be the number of elements of a set T. Set

$$\alpha(k) = 1 + \sum_{p=0}^{s} \operatorname{ord}(\Phi_p^k). \tag{10}$$

Let S_m, $1 \leqq m \leqq \alpha(k)$, be a collection of disjoint P-sets with the further property that if $g_1 \in S_{m_1}$ and $g_2 \in S_{m_2}$, then $g_1g_2^{-1}$ is not a member of any subgroup K such that a coset aK is contained in U_k. That this can be done may be seen by first choosing S_1, then choosing S_2 disjoint from $S_1 + K$ and so on. Again we assume that S_m are P-sets in the strong sense that there are no exceptional values. Let $f \in E$ and $g_m \in S_m$ so that $fg_m \in E$. To each fg_m there exists $p(m)$ satisfying property B. By (10) there are two values m_1 and m_2 such that the function ϕ_1 and ϕ_2 assigned by B are the same. Put $g = g_{m_1}^{-1}g_{m_2}$. We now have two cases. Case I occurs when g is of infinite order, Case II when g is of finite order. In Case I we define ψ exactly as

previously described. We remark again that only finitely many functions ψ/f can occur. In this case we assign to f and $k+1$ under property B, the integer $p(m_1)+1$ and the function ψ/f. This means of course that ψ/f must be placed in the set $\Phi_{p(m_1)+1}^{k+1}$. Since $\int\psi\,d\mu = 1+[p(m_1)+1]\delta(M)$ it is clear that $p(m_1)+1 \leqq s$. If the cosets $a_0K_0, a_1K_1, \cdots, a_{k-p}K_{k-p}$ are assigned to fg_{m_1} and k under condition C, then we assign $a_0g_{m_1}K_0, a_1g_{m_1}K_1, \cdots, a_{k-p}g_{m_1}K_{k-p}$ to f and $k+1$. This is possible if we define U_{k+1} to include U_k, since we already have $U_j \subset U_{j+1}$ for $j \leqq k-1$. It is now clear that conditions A, B, and C are completely fulfilled in this manner. In Case II g is of finite order. Let aK be assigned to fg_{m_1} and k under A. The set $\{Kg^n\}$ is a finite group properly containing K since $g \notin K$. We now distinguish two possibilities. If $\{fg_{m_1}aKg^n\}$ is contained in E, then if ϕ and p in condition B are assigned to fg_{m_1} and k, we assign $g_{m_1}\phi$ and p to f and $k+1$. To fulfill C we must find a chain of length $k-p+2$ satisfying C. But if we assign $g_{m_1}aKg^n$ to f and $k+1$ under A, clearly $\{g_{m_1}aKg^n\}, \{g_{m_1}aK\}, \{g_{m_1}a_1K_1\}, \cdots, \{g_{m_1}a_{k-p}K_{k-p}\}$ is such a chain. We thus adjoin $\{g_{m_1}aKg^n\}$ to U_{k+1}. The number of such cosets we must adjoin is clearly finite. The second possibility occurs if $\{fg_{m_1}aKg^n\}$ is not contained in E. This means that for some b in $\{g_{m_1}aKg^n\}$, $fb \in E$ but $fbg \notin E$. The number of such b which can occur is clearly finite. We now apply Lemma 6, where the roles of ϕ, g, and h are played by $fg_{m_1}\phi_1, g_{m_2}g_{m_1}^{-1} = g$, and fg respectively. As before $1+p\delta(M)$ now plays the role of M. Thus there is a ψ, where

$$\psi = f(\alpha_1 g_{m_1}\phi_1 + \alpha_2 g_{m_2}\phi_1 + \alpha_3 b + \alpha_4 bg)$$

and $|\psi| \leqq 1$, $\int\psi\,d\mu = 1+(p+1)\delta(M)$. In this case now we assign ψ/f to f and $k+1$. In C, we assign $g_{m_1}aK, g_{m_1}a_1K_1, \cdots, g_{m_1}a_{k-p}K_{k-p}$ to f and $k+1$. We adjoin ψ/f to $\Phi_{p(m_1)+1}^{k+1}$. Clearly we perform only finitely many adjunctions. Our construction of Φ_p^k is now completed if we further demand $\Phi_p^k \subset \Phi_p^{k+1}$.

Now, as k tends to infinity for f fixed, it is clear that there are longer and longer chains occurring in condition C. For $k > s$, since $p \leqq s$, at least one of the a_jK_j belongs to U_s. Since each U_k is finite the usual selection argument shows that there is an infinite chain $\bar{K}_1 \subset \bar{K}_2 \subset \cdots$ such that for every j, there is a coset $a_j\bar{K}_j$, such that $fa_j\bar{K}_j$ is contained in E. Now by Lemma 7, this implies that μ has mass on a subgroup of infinite index which is a contradiction. Thus the Main Lemma is entirely proved.

Note. Since this paper was written Professor Davenport has shown that

Theorem 2 holds with the exponent 1/4. This was accomplished by improving the combinatorial Lemma 4 so that instead of $(2r^2)^{2s^2}$ appearing there, something of the form $(2r^2)^{cs}$ takes its place. Also he has explicitly given values for the constants α, β_j, γ_{ij} in the corollary to Lemma 3 and proved directly that they satisfy the required inequalities.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] G. H. Hardy and J. E. Littlewood, "A new proof of a theorem on rearrangements," *Journal of the London Mathematical Society*, vol. 23 (1948), pp. 163-168.

[2] S. Helgason, "Some problems in the theory of almost periodic functions," *Mathematica Scandinavica*, vol. 3 (1955), pp. 49-67.

[3] H. Helson, "Note on harmonic functions," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 686-691.

[4] L. Loomis, *Abstract Harmonic Analysis*, New York, 1953.

[5] W. Rudin, "Idempotent measures on abelian groups," *Pacific Journal of Mathematics*, vol. 9 (1959), pp. 195-209.

[6] R. Salem, "On a problem of Littlewood," *American Journal of Mathematics*, vol. 77 (1955), pp. 535-540.

[7] A. Zygmund, *Trigonometrical Series*, Warsaw, 1935.

ON HOMOMORPHISMS OF GROUP ALGEBRAS.*

By PAUL J. COHEN.

Introduction. There have been several papers written on the subject of determining all homomorphisms, and more particularly, isomorphisms of group algebras of locally compact abelian groups. Except in certain special cases there does not seem to be much known. One case which has been treated is the case of two groups G and H, where H has a connected dual group $\hat{H}$. In this case it has been shown that the only homomorphisms of $L^1(G)$ into $L^1(H)$ are essentially those induced by homomorphisms of G into H [1]. This result was proved in the case where H is the real line and by means of the structure theory of locally compact abelian groups, extended to the more general case. The crucial point in the proof seems to be the obvious fact that a Fourier-Stieltjes transform taking only the values zero and one must either be identically zero or identically one. Equivalently we may say that the only idempotent measures on H are the zero measure, and Haar measure of the identity subgroup. Another case, that in which H is the circle group, has been solved in [5], [6]. In this case too, the complete analysis of idempotent measures on the circle achieved in [4], was very heavily used. The author in a previous paper [2], has determined all the idempotent measures on locally compact groups. Thus, it seems reasonable that one should now be able to completely solve the homomorphism problem.

In this paper, we shall do exactly that. A very simple, but hitherto unnoticed, relationship between the homomorphism problem and idempotent measures is established in the case of compact G and H. Then a passage to the Bohr compactifications of the groups in question yields the general result. There are certain technical complications which appear, some of which are standard, such as convolutions with approximate identities, which we hope will not confuse the reader. It is perhaps unnecessary to add that at all times the reader should bear in mind the concrete examples of Fourier series and Fourier integrals to better understand what is happening. In the case of Fourier series, our problem is precisely one of determining which mappings of Fourier coefficients in m-variables into coefficients in n-variables, send Fourier coefficients into Fourier coefficients. More specifically, let π be a

* Received January 8, 1959.

map from a subset J of the set of all m-tuples of integers, into the set of all n-tuples of integers. We then ask, under what conditions on π and J, will it be true that whenever

$$\sum_{a} c_a e^{i(a_1x_1+\cdots+a_nx_n)}, \qquad a=(a_1,\cdots,a_n), \tag{1}$$

is a Fourier series in n-variables, will it follow that

$$\sum_{b\in J} c_{\pi(b)} e^{i(b_1x_1+\cdots+b_nx_n)}, \qquad b=(b_1,\cdots,b_m), \tag{2}$$

is a Fourier series in m-variables? The answer to this question is contained in Theorem 1 below, though it is difficult to express the answer any more simply in this special case than in the general situation. Suffice it to say that J is composed from linear subspaces of n-tuples by means of Boolean operations, and π is a piecewise linear map.

1. We begin by reviewing certain standard notations and definitions If G is a locally compact abeliean group, and dm is Haar measure on G, $L^1(G)$ is defined to be the space of all measurable functions $f(x)$ on G, such that

$$\|f\| = \int_G |f(x)|\, dm(x) \tag{3}$$

exists and is finite. $L^1(G)$ is a Banach algebra where the product, or as it is customarily called, the convolution of two functions f and g is defined as

$$(f*g)(x) = \int_G f(t)g(x-t)\,dm(t). \tag{4}$$

The Banach algebra of measures on G, denoted by $M(G)$, consists of all finite Borel measures μ, where $\|\mu\|$ is defined as the total variation of μ and the convolution $\mu*\nu$ is defined as

$$(\mu*\nu)(E) = \int\int_{x+y\in E} d\mu(x)\,d\nu(y). \tag{5}$$

The maximal ideals of $L^1(G)$ with their topology can be identified with $\hat{G}$, the dual group of G. If χ is an element of $\hat{G}$, we use (χ, t) to denote χ applied to t, where t is an element of G. Then the map

$$\alpha: f(x) \to \int_G f(t)(\chi, t)\,dm(t) \tag{6}$$

is a homomorphism of $L^1(G)$ into the complex numbers and in this manner

one defines the correspondence between elements of $\hat{G}$ and maximal ideals of $L^1(G)$. The map α extends to $M(G)$, namely,

$$\alpha: \mu \rightarrow \int_G (\chi, t)\, d\mu(t). \tag{7}$$

Since $L^1(G)$ is a closed ideal in $M(G)$, namely the ideal of absolutely continuous measures, and since $L^1(G)$ is semi-simple, it is clear that one does not obtain all the maximal ideals of $M(G)$ by means of (7). Equations (6) and (7) define function on $\hat{G}$ denoted by $\hat{f}(\chi)$ and $\hat{\mu}(\chi)$ respectively, which are known as the Fourier transforms of f and μ.

Let ϕ now be a homomorphism of $L^1(G)$ into $M(H)$. It is well known that ϕ is continuous [5]. For any homomorphism α, of $M(H)$ into the complexes, the map $f \rightarrow \alpha(\phi(f))$ is either a non-trivial homomorphism of $L^1(G)$ into the complexes, or is identically zero. In particular, we thus have a map ϕ_* from $\hat{H}$ into the set $\{\hat{G}, 0\}$, consisting of the group $\hat{G}$ with the symbol 0 adjoined. If we topologize $\{\hat{G}, 0\}$ so that 0 is the point at infinity in the one point compactification of $\hat{G}$, if $\hat{G}$ is not compact, while if $\hat{G}$ is compact we adjoin 0 as an isolated point, it will then follow that ϕ_* is a continuous map. Since a measure in $M(H)$ is determined by its Fourier transform, it follows that ϕ_* completely determines ϕ. Our task is then to characterize the maps ϕ_*. The main result is the following theorem:

THEOREM 1. *Let G and H be locally compact abelian groups, ϕ a homomorphism of $L^1(G)$ into $M(H)$, ϕ_* the induced map of $\hat{H}$ into $\{\hat{G}, 0\}$. Then there are a finite number of sets K_i, which are cosets of open subgroups of $\hat{H}$, and continuous maps ψ_i: $K_i \rightarrow \hat{G}$, such that*

$$\psi_i(x + y - z) = \psi_i(x) + \psi_i(y) - \psi_i(z) \tag{8}$$

for all x, y and z in K_i, with the following property: There is a decomposition of $\hat{H}$ into the disjoint union of sets S_j, each lying in the Boolean ring generated by the sets K_i, such that on each S_j, ϕ_* is either identically zero or agrees with some ψ_i, where $S_j \subset K_i$.

Conversely, for any such map of $\hat{H}$ into $\{\hat{G}, 0\}$, there is a homomorphism of $L^1(G)$ into $M(H)$ which induces it. The map ϕ carries $L^1(G)$ into $L^1(H)$, if and only if ϕ_*^{-1} of every compact subset of $\hat{G}$ is compact.

In the course of the proof, we shall need the characterization of idempotent measures referred to above. The form in which we state the theorem differs slightly from the version in [2]. This is because in that paper we considered idempotent measures on compact groups only, having previously remarked that

any idempotent measure has its support on a compact subgroup. Bearing this in mind, it is easy to see that the following formulation is equivalent to the one given in [2].

THEOREM 2. *If μ is a measure on a locally compact abelian group G, such that $\mu * \mu = \mu$, then $\hat{\mu}$ is the characteristic function of some set E, the sets E which can occur being precisely those which lie in the Boolean ring generated by cosets of open subgroups of $\hat{G}$.*

We shall first prove the second half of Theorem 1. Let us first treat the simplest case, namely, we assume that we have a continuous homomorphism ϕ_* of $\hat{L}$ into $\hat{G}$. We would like to show that there is a homomorphism ϕ of $L^1(G)$ into $M(H)$ inducing ϕ_*. We observe first that there is a continuous homomorphism $\bar{\phi}$ of G into H such that

$$(\chi, \bar{\phi}(g)) = (\phi_*(\chi), g) \tag{9}$$

for all χ and g in $\hat{L}$ and $\hat{G}$ respectively. The map $\bar{\phi}$ induces a map ϕ of $M(G)$ into $M(H)$ by means of the formula,

$$(\phi(\mu))(E) = \mu(\bar{\phi}^{-1}(E)). \tag{10}$$

For any Borel set E in H, we have

$$\begin{aligned}[\phi(\mu) * \phi(\nu)](E) &= \int\int_{x+y \in E} d\phi(\mu)(x)\, d\phi(\mu)(y) \\ &= \int\int_{\bar{\phi}(g_1)+\bar{\phi}(g_2) \in E} d\mu(x)\, d\nu(y)\end{aligned} \tag{11}$$

by (10). Since $\bar{\phi}$ is a homomorphism, the last integral in (11) is equal to

$$\int\int_{g_1+g_2 \in \bar{\phi}^{-1}(E)} d\mu(x)\, d\nu(y) = [\phi(\mu * \nu)](E). \tag{12}$$

Thus ϕ is a homomorphism of $M(G)$ into $M(H)$. For a fixed choice dm, of Haar measure on G, $L^1(G)$ is canonically imbedded in $M(G)$, so that ϕ is also a map of $L^1(G)$ into $M(H)$. To see that ϕ induces ϕ_* we need only verify that

$$\int_G (\phi_*(\chi), t)\, d\mu(t) = \int_H (\chi, t)\, d\phi(\mu)(t). \tag{13}$$

But, the left side of (13) is equal to

$$\int_G (\chi, \bar{\phi}(t))\, d\mu(t) = \int_H (\chi, t)\, d\phi(\mu)(t) \tag{14}$$

by (10), so that ϕ does induce ϕ_*.

Now, we notice that if a homomorphism ϕ induces the map ϕ_*, then clearly for any character χ in $\hat{G}$, the map $\chi\phi_*$ is induced by the map $\mu \rightarrow \phi(\chi\mu)$, as we see from the formula

$$(15) \qquad \int_G (t, \phi_*(\alpha))\, d\mu(t) = \int_H (t, \alpha)\, d\phi(\mu)(t)$$

for all α in $\hat{H}$. Similarly we can see that the map $\phi_*' \colon \alpha \rightarrow \phi_*(\alpha \cdot \chi)$, where α and χ are in $\hat{H}$, is induced by the homomorphism $\mu \rightarrow \chi \cdot \phi(\mu)$. We can thus say that the set of all ϕ_* is invariant under the operation of translation, either in the image or domain space. Now, if ψ is a map from a coset K of some open subgroup L of $\hat{H}$ into $\hat{G}$, such that

$$(16) \qquad \psi(x + y - z) = \psi(x) + \psi(y) - \psi(z)$$

for x, y and z in K, *then for suitable* a in $\hat{G}$, b in $\hat{H}$, we have

$$\psi(x) = a + \phi_*(x + b),$$

where ϕ_* is a homomorphism of L into $\hat{G}$. If we can show that ϕ_* is induced by a homomorphism of $L^1(G)$ into $M(H)$ it will then follow that ψ is likewise induced by such a homomorphism. We have just seen that this is so if L is all of $\hat{H}$. If L is an open subgroup of $\hat{H}$, then L is the dual group of H/H_1, where H_1 is a compact subgroup of H. It thus follows there is a map ϕ' of $L^1(G)$ into $M(H/H_1)$ such that

$$(17) \qquad \int_G (t, \phi_*(\alpha))\, d\mu(t) = \int_{H/H_1} (t, \alpha)\, d\phi'(\phi)(t)$$

for α in L. We will now show that there is a homomorphism σ of $M(H/H_1)$ into $M(H)$, such that for all μ in $M(H/H_1)$

$$(18) \qquad \int_H (\chi, t)\, d\sigma(\mu) = \int_{H/H_1} (\chi, t)\, d\mu$$

if χ is in L and zero otherwise. Composing the map ϕ' with σ will then show that ϕ_* is indeed induced by a homomorphism from $L^1(G)$ into $M(H)$. Such a map σ is given by the formula

$$(19) \qquad \int_H f\, d\sigma(\mu) = \int_{H/H_1} d\mu(x) \int_{H_1} f(t - x)\, dm(t),$$

where $m(t)$ is Haar measure on H_1, normalized to have total mass one, and f is any continuous function on H.

Let us consider the most general case of the second part of Theorem 1. By Theorem 2, to each set S_j there exists a measure μ_j on H, such that $\hat{\mu}_j$ is

the characteristic function of S_j. If S_j is contained in K_i, and ϕ' is a homomorphism of $L^1(G)$ into $M(H)$ such that ϕ'_* agrees with ψ_i on K_i, and is zero elsewhere, then it is clear that the map $\phi: \mu \to \mu_j * \phi'(\mu)$ induces a map ϕ_* which agrees with ψ_i on S_j, and is zero elsewhere. By summing all such maps for each j, we see that the second half of Theorem 1, without the assertion concerning when $L^1(G)$ is mapped into $L^1(H)$, is proved.

2. Let ϕ now be a homomorphism from $L^1(G)$ into $M(H)$. We shall first prove that ϕ can be extended to a map of $M(G)$ into $M(H)$, and give conditions when this extension is unique. These elementary facts are essentially contained in [3] and [5], but we reprove them here for the sake of completeness. Also, we do not use directed sets as in [3]. Assume first that ϕ can be extended to a map of $M(G)$ into $M(H)$. Then for each χ in $\hat{E}$, the map

$$\mu \to \int_H \chi \, d\phi(\mu) \tag{20}$$

is a homomorphism of $M(G)$ into the complexes. On $L^1(G)$ this map coincides with the map

$$\mu \to \int_G \phi_*(\chi) \, d\mu, \tag{21}$$

where μ is an absolutely continuous measure. If $\phi_*(\chi)$ is not zero the maps (20) and (21) must agree for all μ. This follows from the general fact that a homomorphism θ from a ring R into a field, which is not identically zero on an ideal I, is determined by θ restricted to I. For, if a is an element of I such that $\theta(a) \neq 0$, for all x in R we have

$$\theta(x) = \theta(ax)/\theta(a), \tag{22}$$

where ax also belongs to I. Thus, if ϕ_* is never zero, ϕ extends uniquely to $M(G)$, if at all. Now, if we can show that for each μ in $M(G)$, there exists $\phi(\mu)$ in $M(H)$, such that

$$\int_H \chi \, d\phi(\mu) = \int_G \phi_*(\chi) \, d\mu \tag{23}$$

for all χ in $\hat{E}$, then clearly ϕ will then be a homomorphism of $M(G)$ into $M(H)$ extending our original homomorphism and inducing the same ϕ_*. If we set

$$\lambda(\chi) = \hat{\mu}(\phi_*(\chi)), \tag{24}$$

we are asking whether or not λ is the Fourier transform of a measure. For each continuous function g on H, vanishing at infinity, whose Fourier transform has compact support, consider the map β,

$$\beta: g \to \int_{\hat{H}} \hat{g}\lambda \, dm. \tag{25}$$

It is known that the set of all such g are dense in the space of all continuous functions vanishing at infinity. If we show that β is a bounded functional, it will then follow that for some measure ν on H, we have

$$\int_H g \, d\nu = \int_{\hat{H}} \hat{g}\lambda \, dm. \tag{26}$$

On the other hand, we have

$$\int_H g \, d\nu = \int_{\hat{H}} \hat{g}\hat{\nu} \, dm. \tag{27}$$

Since, the set of all $\hat{g}$ is dense in the space of continuous functions on $\hat{H}$ vanishing at infinity, we will have $\lambda = \nu$. Therefore, we need only show that β is a bounded functional. For a fixed g, let $\hat{g}$ vanish outside S, where S is compact. For $\epsilon > 0$, let k be a function in $L^1(G)$ such that $\| k \| = 1$, and $| \hat{k} - 1 | < \epsilon$ on $\phi_*(S)$. Now, we have

$$\widehat{\phi(k)}\lambda = \hat{k}(\phi_*)\hat{\mu}(\phi_*) = (\widehat{k*\mu})(\phi_*) = \widehat{\phi(k*\mu)}. \tag{28}$$

If C is a bound for ϕ, that is,

$$\| \phi(f) \| \leqq C \| f \|, \tag{29}$$

then, for all f in $L^1(G)$ we have

$$\begin{aligned} \left| \int_{\hat{H}} \hat{g}\widehat{\phi(k)}\lambda \, dm \right| &= \left| \int_{\hat{H}} \hat{g}\widehat{\phi(k*\mu)} \, dm \right| \\ &= \left| \int_H g\phi(k*\mu dm \right|. \end{aligned} \tag{30}$$

Hence the left side of (30) is bounded by

$$\| g \|_\infty \| \phi(g*\mu) \| \leqq C \| g \|_\infty \| \mu \|. \tag{31}$$

Now as ϵ tends to zero, the left side of (30) approaches the right side of (25). Thus ν exists and $\| \nu \| \leqq C \| \mu \|$.

If ϕ_* maps any element of $\hat{H}$ into zero, it will follow from Theorem 1 that the set of all such elements is an open set in $\hat{H}$, whose characteristic

function is the Fourier transform of an idempotent measure μ_0. If ϕ is the extended map of $M(G)$ into $M(H)$, which induces ϕ_*, which was shown to exist above, and if ω is an arbitrary homomorphism of $M(G)$ into the complexes which vanish on $L^1(G)$, of which there are many, then it is clear that the homomorphism

$$\phi' : \mu \rightarrow \phi(\mu) + \omega(\mu)\mu_0 \tag{32}$$

is a different extension of the original ϕ. Since we have quoted Theorem 1 in the proof of this last point, we shall not use this observation in the sequel. Summarizing, we have

THEOREM 3. *If ϕ is a homomorphism of $L^1(G)$ into $M(H)$, ϕ can be extended to a homomorphism of $M(G)$ into $M(H)$, by means of the formula*

$$\widehat{\phi(\mu)} = \hat{\mu}(\phi_*). \tag{33}$$

The extended ϕ will have the same bound as ϕ itself. This will be the unique extension if and only if ϕ_* never takes the value zero, or G is discrete.

3. In this section we prove the fundamental lemma.

LEMMA. *Let G and H be compact abelian groups, ϕ a homomorphism of $L^1(G)$ into $M(H)$. Let S denote the subset of $\hat{G} \times \hat{H}$ consisting of all pairs $(\phi_*(h), h)$, where h runs through all members of $\hat{H}$ such that $\phi_*(h)$ is not zero. Then the characteristic function of S is the Fourier transform of an idempotent measure on $G \times H$.*

Proof. Let k be a function on H, such that $\| k \| = 1$, and $\hat{k}$ takes non-zero values only finitely often. Thus k is a finite exponential sum on H. We shall have occasion to regard k as a function on $G \times H$, and $\hat{k}$ as a function on $\hat{G} \times \hat{H}$. Of course, now $\hat{k}$ may take non-zero values infinitely often. As in (29) let C be a bound for the map ϕ. We shall first show that if λ is the characteristic function of S, $\lambda\hat{k}$ is the Fourier transform of a measure of norm not greater than C. Since $\hat{k}$ can be made arbitrarily close to one on given finite sets, it then will follow, as in the previous section, that λ itself is the Fourier transform of a measure of norm not exceeding C. We have that $\lambda\hat{k}$ is the Fourier transform of the following function on $G \times H$,

$$\phi(g, h) = \sum_{\chi \in \hat{H}} \hat{k}(\chi)\, \overline{(\phi_*(\chi), g)}\, (\bar{\chi}, h), \tag{34}$$

where g and h are in G and H respectively. This is of course only a finite sum. For a fixed g, $\overline{(\phi_*(\chi), g)}$ is the Fourier transform of $\phi(\mu_{-g})$, where ϕ

here denotes the extension of our original ϕ to the measure algebra $M(G)$, and μ_{-g} is a point mass of one at $-g$. Hence, for fixed g, $\phi(g,h)$ is the convolution of k and $\phi(\mu_{-g})$. Thus,

$$\int_H |\phi(g,h)|\,dm(h) \leqq \|k\| \cdot \|\phi(\mu_{-g})\| \leqq C. \tag{35}$$

Since this holds for each g and the total measure of G under Haar measure is one, we have

$$\|\phi(g,h)\| \leqq C \tag{36}$$

and the lemma is proved.

We shall say that a subset of $\hat{G} \times \hat{H}$ is a graph, if for each h in $\hat{H}$ there is at most one g in $\hat{G}$, such that (g,h) is in that set. By Theorem 2, S is in the Boolean ring generated by cosets of subgroups of $\hat{G} \times \hat{H}$. S is of course also a graph, and we shall now show that S is in the Boolean ring generated by those cosets of subgroups of $\hat{G} \times \hat{H}$, which are also graphs. We need a simple lemma.

LEMMA. *An abelian group cannot be the union of a finite number of cosets of subgroups, each of infinite index.*

Proof. Let G denote the group, and assume that G is the union of cosets $a^i{}_j + H_i$, where H_i are distinct subgroups, each of infinite index. If for any pair i_1 and i_2, we have that $[H_{i_1}: H_{i_1} \cap H_{i_2}]$ is finite and distinct from one, then it is clear that every coset of H_{i_1} is a finite union of cosets of $H_{i_1} \cap H_{i_2}$. Hence we can replace H_{i_1} by $H_{i_1} \cap H_{i_2}$ in our original list of subgroups H_i. After a finite number of such replacements, it is clear that without loss of generality, we may assume that for all i_1 and i_2, $[H_{i_1}: H_{i_1} \cap H_{i_2}]$ is either one or infinite. Let K now be any coset of H_1 which does not appear in our finite collection. The intersection of K with a coset of H_i, $i > 1$, is clearly a coset of the subgroup $H_1 \cap H_i$. Thus K must be the union of a finite number of cosets of $H_1 \cap H_i$, $i > 1$. Equivalently, H_1 must be such a union. Now if we assume by induction that the lemma is true if no more than k distinct subgroups H_i appear, we have clearly proved it for the case $k+1$. Hence the lemma follows by induction.

Let S now be in the Boolean ring generated by cosets K_i of subgroups H_i of $\hat{G} \times \hat{H}$, $1 \leqq i \leqq n$. We assume for the sake of convenience that the group $\hat{G} \times \hat{H}$ itself appears among the K_i. Then S is the union of sets of the form

$$P = (\bigcap_{j=1}^{r} K_{i_j}) \cap (\bigcap_{j=r+1}^{n} CK_{i_j}), \tag{37}$$

where CK_{i_j} denotes the complement of K_{i_j}. By a finite number of replacements as above, we can assume that $[H_i\colon H_i \cap H_j]$ is either one or infinite. The set P can also be written as

$$P = L \cap \left(\bigcap_{j=r+1}^{n} CM_j \right), \tag{38}$$

where $L = \bigcap_{j=1}^{r} K_{i_j}$, and $M_j = L \cap K_{i_j}$. Now M_j is contained in L, and if L is a coset of L', while M_j is a coset of M'_j, $[L'\colon M'_j]$ is either one or infinite. Since S is a graph, and S contains P, P is also a graph. We shall now show that L is a graph. Assume that L' did contain an element of the form $(g, 0)$, where g belongs to $\hat{G}$ and is not zero. Then for (g_1, h_1) in L, either (g_1, h_1) or $(g_1 + g, h_1)$ is in $\bigcup_{j=r+1}^{n} M_j$, since otherwise P would not be a graph. Hence L is the union of M_j and $M_j + (g, 0)$, which violates the lemma. Thus L and hence M_j are graphs. Since the projection on $\hat{H}$ is in the Boolean ring generated by cosets of subgroups of $\hat{H}$, and L is the graph of a linear function, we see that the first half of Theorem 1 in the case of compact G and H is proved.

4. We now pass to the case where G and H are not necessarily compact. As described in the introduction we shall have occasion to pass to the Bohr compactifications $\bar{G}$ and $\bar{H}$ of G and H respectively.

LEMMA. *Let ϕ be a homomorphism from $L^1(G)$ into $M(H)$ inducing $\phi_*\colon \hat{H} \to \{\hat{G}, 0\}$. If we then consider $\hat{G}$ and $\hat{H}$ in the discrete topology, ϕ_* satisfies the conditions of Theorem* 1.

Proof. We need only show that there is a homomorphism ϕ' of $L^1(\bar{G})$ into $M(\bar{H})$ inducing ϕ_* on $\hat{H}$ into $\hat{G}$. We recall here that $\bar{G}$ and $\bar{H}$ are the dual groups of $\hat{G}$ and $\hat{H}$ taken with the discrete topology. If we compose the map of $M(H)$ into $M(\bar{H})$ induced by the canonical imbedding $\pi\colon H \to \bar{H}$ with ϕ, we obtain a homomorphism λ of $L^1(G)$ into $M(\bar{H})$ which induces ϕ_* on $\hat{H}$ into $\{\hat{G}, 0\}$.

As in § 2, by convoluting with an approximate identity, it only remains to show that for any μ in $M(\bar{G})$, the function $\hat{\mu}(\phi_*)$ on $\hat{H}$ can be approximated arbitrarily well on finite subsets by functions of the form $\hat{\nu}$, where $\| \nu \|$ is bounded. For any finite subset of $\hat{G}$, one can find a measure μ' in $M(G)$ such that $\hat{\mu}'$ and $\hat{\mu}$ differ by arbitrarily little on that subset and $\| \mu' \| \leqq \| \mu \|$. This is true because G is dense in $\bar{G}$. Since

$$\| \lambda(\mu') \| \leqq C \| \mu' \| \leqq C \| \mu \|,$$

$\lambda(\mu')$ will serve as the ν above, and we have proved the lemma.

The rest of this section will now be devoted to the task of showing that ϕ_* does satisfy the conditions of Theorem 1, with the original topology on $\hat{G}$ and $\hat{H}$. We shall deduce this from the preceding lemma and the fact that ϕ_* is continuous. We shall first show that if T denotes the subset of $\hat{H}$ on which ϕ_* is not zero, then T is an open and closed set. We know that S, the graph of ϕ_*, is a union of sets of the form (38), where L and M_j are graphs, and if L is a coset of L', while M_j are cosets of M'_j, L contains M_j, and $[L' : M'_j]$ is infinite. It follows then that T is the union of sets of the form

$$\tilde{P} = \tilde{L} \cap (\bigcap_{j=1}^{r} C\tilde{M}_j), \tag{39}$$

where $\tilde{L}$ and $\tilde{M}_j$ are the projections on $\hat{H}$ of L and M_j respectively. Fuuthermore on $\tilde{P}$, ϕ_* agrees with a map ψ of $\tilde{L}$ into $\hat{G}$, where

$$\psi(x + y - z) = \psi(x) + \psi(y) - \psi(z) \tag{40}$$

for x, y and z in $\tilde{L}$. By the lemma of § 3, $\tilde{L}$ is not contained in the union of a finite number of translates of the $\tilde{M}_j$. It follows easily that there are $r + 1$ elements in $\tilde{P}$, $a_1, \cdots, a_{r+1}$ such that for $m \neq n$, $a_m - a_n$ is never in the subgroup $\tilde{M}'_j$ of which $\tilde{M}_j$ is a coset. Hence, if x and y are in $\tilde{L}$, if for all a_m, we had that $a_m + x - y$ were not in $\tilde{P}$, for some m and n, $a_m + x - y$ and $a_n + x - y$ would be in the same $\tilde{M}_j$. This would imply that $a_m - a_n$ is in $\tilde{M}'_j$. Thus for x and y in $\tilde{L}$, there is some m such that $a_m + x - y$ is in $\tilde{P}$. If U is a neighborhood of the identity in $\hat{G}$, we can find open sets V_m containing a_m, such that z_1 and z_2 in $V_m \cap \tilde{P}$ implies that $\phi_*(z_1) - \phi_*(z_2)$ is in U. Now

$$\psi(x) - \psi(y) = \phi_*(a_m + x - y) - \phi_*(a_m). \tag{41}$$

Hence for some neighborhood V of the identity, if x and y are in $\tilde{L}$, and $x - y$ lies in V, then $\psi(x) - \psi(y)$ is in U. Hence ψ is uniformly continuous on $\tilde{L}$ and hence on $\tilde{P}$. Thus the closure of $\tilde{P}$ is also contained in T, since zero is either the point at infinity on $\hat{G}$ or is isolated. Therefore, T is the union of a finite number of closed sets, and so is closed. On the other hand, T is trivially open, so that T is both open and closed. In particular, S is a closed set.

To avoid the use of subscripts, we introduce the notation $f(A, x)$ for the characteristic function of the set A. We have then the following lemma.

LEMMA. *Let K_i^j be a finite collection of cosets of subgroups H_i of a group G. Then if*

$$f(x) = \sum c_{ij} f(K_i^j, x) \tag{42}$$

for some constants c_{ij}, and B_k are the disjoint sets on which $f(x)$ takes its finite number of values, then there are subgroups H'_i lying in the Boolean ring generated by B_k and their translates, such that for some cosets $K'_i{}^j$ of H'_i each B_k is in the Boolean ring generated by $K'_i{}^j$.

Proof. As before we may assume that $[H_i : H_i \cap H_j]$ is either always one or infinity. Under this assumption we shall prove that if

$$f_i(x) = \sum_j c_{ij} f(K_i{}^j, x) \tag{43}$$

takes its finite number of values on sets C_k, each of which is of course a union of cosets of H_i, then each C_k is a union of cosets of some subgroup H'_i, which, in turn, lie in the Boolean ring generated by all the sets B_k and their translates. Also $[H'_i : H_i]$ is finite. The proof proceeds by induction on the number of distinct groups H_i which occur. Assume first that there is only one $H_i = H$. It clearly suffices to prove it in the case where c_{ij} are all one. Let H' be the group of all a such that $f(x+a) = f(x)$. H' contains H and we have

$$f(x) = \sum_{j=1}^{s} f(H' + b_j, x), \tag{44}$$

where b_j are elements of G. By translating f we can assume that $b_1 = 0$. We prove, by induction on s, that H' is in the Boolean ring generated by the support of f_i and its translates. The case $s = 1$ is trivial. The union of all the cosets $H' + b_j$ does not form a group, since otherwise the definition of H' would be contradicted. Hence for some b_j, $f(x - b_j) - f(x)$ is not identically zero.

But

$$f(x - b_j) - f(x) = g_1(x) - g_2(x), \tag{45}$$

where g_1 and g_2 are characteristic functions of unions of less than s cosets of H'. By induction it follows that the group H'_1 which leaves the support of g_1 invariant under translation is in the Boolean ring we would like it to be in. Also by considering $f - g_1$, the subgroup H'_2 leaving it invariant is also in that Boolean ring. Since H' is the intersection of H'_1 and H'_2 the result follows in this case.

If more than one H_i occurs, assume that H_1 is such that it is not contained in any other H_i. Then H_1 is not contained in the union of any finite number of cosets of H_i, $i \neq 1$. Let a_1 and a_2 be two members of H_1 such that the cosets $K_2{}^j$, $K_2{}^j + a_1$, $K_2{}^j + a_2$ are all distinct. Now the functions

$f(x)-f(x-a_1)$ and $f(x)-f(x-a_2)$ involve fewer distinct H_i than f, since they do not involve H_1, and so

$$\sum c_{2j}f(K_2{}^j,x)-\sum c_{2j}f(K_2{}^j+a_1,x) \tag{46}$$

and

$$\sum c_{2j}f(K_2{}^j,x)-\sum c_{2j}f(K_2{}^j+a_2,x) \tag{47}$$

take their distinct values on unions of cosets H''_2 and H'''_2 respectively, both of which contain H_2, and which lie in the desired Boolean ring. From the choice of a_1 and a_2 it follows that $f_2(x)$ takes its distinct values on unions of cosets of $H'_2=H''_2\cap H'''_2$. Similarly for all f_i, $i\neq 1$. Hence the lemma is proved.

We now apply the lemma to the graph S of ϕ_*, which we know is a closed set. We deduce that S is a union of sets P as in (38), where L and M_j are in the Boolean ring generated by closed sets. Such sets we shall call finite Borel sets. We now would like to deduce that the sets $\tilde{L}$ and $\tilde{M}_j$ in (39), which are the projection of L and M_j are measurable subsets of $\hat{E}$. We could quote a general theorem to this effect, but we prefer to prove it directly. As was proved above, we know that L and M_j are the graphs of uniformly continuous functions. Now, L is the union of a finite number of sets, each of which is an intersection $U\cap F$, where U is open and F is closed. Of course, $U\cap F$ is also the graph of a uniformly continuous function. Hence, without loss of generality we may assume that $L=U\cap F$. We have $L=U\cap \bar{L}$, where $\bar{L}$ is the closure of L. Now by the assumption of uniform continuity $\bar{L}$ is the graph of a continuous function on the closure of $\tilde{L}$. It follows easily that $\tilde{L}$ is the intersection of its closure and an open subset of $\hat{H}$. Hence $\tilde{L}$, and similarly $\tilde{M}_j$, are Borel subsets of $\hat{H}$.

Assume that for a given term $\tilde{P}$, $\tilde{L}$ is an open coset. We need the remark that a subgroup and hence a coset of positive measure, is necessarily open. This is true because it must contain its own set of differences, which is well known to contain a neighborhood of the identity. Let $\tilde{M}_j$, $1\leqq j\leqq s$, be open, while, for $s+1\leqq j\leqq r$, $\tilde{M}_j$ are not open. Then consider the set

$$W=\tilde{L}\cap(\bigcap_{j=1}^{s} C\tilde{M}_j). \tag{48}$$

This set is open. We assert that $\tilde{P}$ is dense in it. Otherwise, $W-\tilde{P}$ contains an open set, but since this set is contained in the union of $\tilde{M}_j$, $s+1\leqq j\leqq r$, each of which is of measure zero, this is impossible. Hence on W, ϕ_* coincides with ψ. Now, the set of all such W covers T. For, the difference

$T - W$ is open and must be contained in the union of a finite number of non-open $\bar{L}$, each of measure zero, which is impossible. Thus we have shown that the first part of Theorem 1 is true.

It only remains to show that ϕ maps $L^1(G)$ into $L^1(H)$ if and only if ϕ_*^{-1} of a compact subset of $\hat{G}$ is always compact. If there is a compact subset C of $\hat{G}$ such that $\phi_*^{-1}(C)$ is not compact, then if f is in $L^1(G)$, such that $\hat{f}$ is greater than or equal to one on C, it is clear that $\hat{f}(\phi_*)$ can not be the Fourier transform of any function in $L^1(H)$. Conversely, if ϕ_* has the above property, and if f is in $L^1(G)$, it is well known that we can find functions f_n in $L^1(G)$ such that $\| f - f_n \|$ tends to zero, and $\hat{f}_n$ vanishes off a compact set. Since $\hat{f}_n(\phi_*)$ vanishes off a compact set, it is clear that $\phi(f_n)$ is in $L^1(H)$. Since $\phi(f_n)$ approaches $\phi(f)$, and $L^1(H)$ is a closed subalgebra of $M(H)$, the result follows.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] A. Beurling and H. Helson, "Fourier-Stieltjes transforms with bounded powers,' *Mathematica Scandinavia*, vol. 1 (1953), pp. 120-126.

[2] P. J. Cohen, "On a conjecture of Littlewood and idempotent measures," *American Journal of Mathematics*, vol. 82 (1960), pp. ——.

[3] H. Helson, "Isomorphisms of Abelian group algebras," *Arkiv för Mathematik*, vol. 2 (1953), pp. 475-487.

[4] ———, "Note on harmonic functions," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 686-691.

[5] W. Rudin, "The automorphisms and endomorphisms of the group algebra of the unit circle," *Acta Mathematica*, vol. 95 (1956), pp. 39-55.

[6] ———, "On isomorphisms of group algebras," *Bulletin of the American Mathematical Society*, vol. 65 (1958), pp. 167-169.

TRIANGULAR OPERATOR ALGEBRAS.*

Fundamentals and Hyperreducible Theory.

By RICHARD V. KADISON and I. M. SINGER.[1]

CONTENTS

Chapter I. Introduction and Preliminaries.

1.1. Introduction. For the past three decades, the theory of self-adjoint operators and self-adjoint operator algebras has undergone a vigorous and moderately successful development. A large share of the credit for this moderate success must be given to the reasonably detailed theory of factors

* Received February 27, 1959.

[1] Alfred P. Sloan Fellows. This work was done in part during the tenure of a National Science Foundation grant.

created by Murray and von Neumann [6], and to its later elaboration, the general theory of von Neumann algebras, by several mathematicians [1]. The von Neumann algebras are special enough so that incisive structural results can be established yet broad enough so that they can be related to the general self-adjoint operator algebras. The limitations which exist at present in the self-adjoint theory seem basically to reside in the special problems about factors which remain unanswered.

It is our hope that the theory we initiate in this tract will be capable of filling an analogous rôle in the study of non-self-adjoint operators and operator algebras. In finite dimensions, the class of operator algebras we study are the triangular operator algebras (the algebras of those matrices relative to given bases in given orderings with 0 entries below the diagonal). In general, the class of algebras we study is characterized by the simple property: $\mathcal{T}^* \cap \mathcal{T}$ is maximal abelian. We call such algebras "triangular"—in finite dimensions, they are subalgebras of the triangular matrices; with a maximality assumption, they are full algebras of triangular matrices. In infinite dimensions one would not expect a "classical" basis to be related to a given maximal triangular algebra in general. The "continuous" as well as the "discrete" appears in infinite dimensions. Even when this is taken into account, however, a large section of the theory must concern itself with maximal triangular algebras to which no ordered basis (in the appropriately general sense) can be said to be associated. Of course the ordering of the basis has a much more critical position in the infinite-dimensional theory than in the finite-dimensional theory (based mainly on the fact that there is just one total-ordering type associated with a given finite set). The hyperreducible maximal triangular algebras (those satisfying certain reducibility conditions) seem to be the correct generalization of the concept of "ordered (orthonormal) basis" in the same sense that maximal abelian (self-adjoint) algebra generalizes the concept of "(orthonormal) basis."

The theory of triangular algebras seems to us to provide the general framework within which the reducibility properties of a bounded operator can be studied—in particular, the questions centered about invariant subspaces and bringing operators to "triangular form." A satisfactory theory of triangular algebras might provide an effective tool for the analysis of infinite-dimensional representations of solvable Lie groups.

In Chapter II, our basic definitions and results as well as examples establishing the existence of various special classes of maximal triangular algebras are presented. Chapter III contains the detailed development of the theory of hyperreducible triangular algebras. These have the position in

triangular theory which the abelian algebras occupy in the study of self-adjoint algebras. It is natural to expect, therefore, that their theory would be most accessible (though the abelian-hyperreducible analogy has very limited applicability). At another time we shall describe the general constructions, with triangular algebras—restriction to a projection, triangular direct sums, and triangular tensor products. These operations are related to the hyper-reducible theory. With the aid of these, some ideal theory for triangular algebras, and the results of [5], an example of a triangular algebra which is not strongly closed can be constructed.

1.2. Preliminaries. Our Hilbert spaces are complex. Our maximal abelian algebras are the self-adjoint ones. All operators are bounded unless otherwise specified. The "multiplication algebra" associated with a measure space is the (maximal abelian) algebra of operators corresponding to multiplications by essentially bounded measurable functions on its L_2 space. "Totally-atomic" maximal abelian algebras are those generated by minimal projections—the "non-atomic" ones are those without minimal projections (atoms). Each maximal abelian algebra is the direct sum of a non-atomic and totally-atomic one. The non-atomic algebras on separable spaces are unitarily equivalent to the multiplication algebra of the unit interval under Lebesgue measure. A separating vector for a maximal abelian algebra is one which is annihilated by no operator in the algebra other than 0.

If $\mathfrak{F}$ is a set of operators and S a set of vectors, we denote by $[\mathfrak{F}S]$ the closed linear space spanned by vectors Tx with T in $\mathfrak{F}$ and x in S, and by $\mathfrak{F}'$ the set of operators commuting with each of those of $\mathfrak{F}$. We often use the same symbol to denote both a projection and its range. Invariance of the range of a projection, E, under an operator, T, is equivalent to $TE = ETE$. Moreover, E is invariant under T and T^* if and only if E commutes with T.

Chapter II. General Theory.

2.1. Basic definitions and notation. A feature of the algebra of $n \times n$ matrices with entries on or above the diagonal relative to a particular basis and a particular ordering of that basis which embodies its characteristic property of having entries on just one side of the diagonal is the fact that its intersection with its adjoint is the set of diagonal matrices, i.e. the maximal abelian algebra associated with the basis. The fact that the algebra contains all such matrices is reflected in its maximality with respects to this intersection property. These considerations lead us to

DEFINITION 2.1.1. *If $\mathcal{M}$ is a factor and $\mathcal{A}$ a maximal abelian (self-adjoint) subalgebra of $\mathcal{M}$; a subalgebra, $\mathcal{J}$, of $\mathcal{M}$ will be said to be "triangular in $\mathcal{M}$" (or simply "triangular," when $\mathcal{M}$ is all bounded operators on the underlying Hilbert space) with diagonal $\mathcal{A}$ when $\mathcal{J} \cap \mathcal{J}^* = \mathcal{A}$. If $\mathcal{J}$ is not a proper subalgebra of another algebra which is triangular in $\mathcal{M}$ we shall say that $\mathcal{J}$ is maximal triangular in $\mathcal{M}$. The projections in $\mathcal{M}$ which are invariant under $\mathcal{J}$ are called "the hulls of $\mathcal{J}$". The intersection of all hulls of $\mathcal{J}$ containing a given projection, E, is called "the hull of E (in $\mathcal{J}$)" and is denoted by "$h_{\mathcal{J}}(E)$." The von Neumann algebra generated by the hulls of $\mathcal{J}$ is called "the core of $\mathcal{J}$."*

Remark 2.1.2. If $\mathcal{J}$ is triangular in $\mathcal{M}$ with diagonal, $\mathcal{A}$, and $\{\mathcal{J}_\alpha\}$ is a family of algebras, containing $\mathcal{J}$, which are triangular in $\mathcal{M}$, and is totally ordered by inclusion; then $\mathcal{J}_0$, the union of the family, is triangular in $\mathcal{M}$ with diagonal $\mathcal{A}$ and contains $\mathcal{J}$. In fact, $\mathcal{J}_\alpha \cap \mathcal{J}_\alpha^*$ is maximal abelian in $\mathcal{M}$, by assumption ($\mathcal{J}_\alpha$ is triangular), and contains $\mathcal{J} \cap \mathcal{J}^* (= \mathcal{A})$, which has also been assumed to be maximal abelian in $\mathcal{M}$. Thus $\mathcal{J}_\alpha \cap \mathcal{J}_\alpha^* = \mathcal{A}$, and $\mathcal{A}$ is the diagonal of each $\mathcal{J}_\alpha$. An operator in $\mathcal{J}_0 \cap \mathcal{J}_0^*$ lies in some $\mathcal{J}_\alpha$ and some $\mathcal{J}_{\alpha'}^*$. Since the family, $\{\mathcal{J}_\alpha\}$, is totally ordered by inclusion, the operator lies in one of $\mathcal{J}_{\alpha'} \cap \mathcal{J}_{\alpha'}^*$, $\mathcal{J}_\alpha \cap \mathcal{J}_\alpha^*$, and thus, in $\mathcal{A}$. Of course, $\mathcal{A}$ is contained in $\mathcal{J}_0 \cap \mathcal{J}_0^*$, so that $\mathcal{A} = \mathcal{J}_0 \cap \mathcal{J}_0^*$; and $\mathcal{J}_0$ is triangular in $\mathcal{M}$ (with diagonal $\mathcal{A}$). Applying Zorn's Lemma, we conclude the existence of a maximal triangular algebra in $\mathcal{M}$ containing $\mathcal{J}$. We have observed, in addition, that when one triangular algebra is contained in another, they have the same diagonal.

Remark 2.1.3. To test that an algebra $\mathcal{J}$ is triangular with diagonal $\mathcal{A}$, it suffices to show that each self-adjoint operator in $\mathcal{J}$ lies in $\mathcal{A}$ and that $\mathcal{A}$ is contained in $\mathcal{J}$. In fact, $\mathcal{J} \cap \mathcal{J}^*$ is a self-adjoint algebra containing each self-adjoint operator in $\mathcal{J}$ and generated (linearly) by these operators.

Remark 2.1.4. If $\mathcal{J}$ is triangular in $\mathcal{M}$ with diagonal $\mathcal{A}$, then each hull of $\mathcal{J}$ is invariant under $\mathcal{J}$, hence under $\mathcal{A}$; hence commutes with and therefore lies in $\mathcal{A}$ (since it lies in $\mathcal{M}$, by assumption, and $\mathcal{A}$ is maximal abelian (self-adjoint) in $\mathcal{M}$). Thus, the hull, $h(E)$, of each projection, E, lies in $\mathcal{A}$; and the core of $\mathcal{J}$ is an (abelian) von Neumann algebra contained in $\mathcal{A}$. It is easy to see that $h(E) = [\mathcal{J}\mathcal{M}'E]$.

In the theory of triangular algebras the core plays the rôle that the center has relative to the theory of self-adjoint operator algebras.

Remark 2.1.5. If $\mathcal{T}$ is a maximal triangular algebra with diagonal $\mathcal{A}$, the center of $\mathcal{T}$ consists of scalar multiples of I. Suppose A is in the center of $\mathcal{T}$, then, since $\mathcal{A}$ is maximal abelian, A lies in $\mathcal{A}$, and, in particular, A is normal. With B in $\mathcal{T}$, $BA = AB$, whence, from Fuglede's Theorem [2], $AB^* = B^*A$, and B commutes with each spectral projection E of A. Thus E and $I - E$ are hulls in $\mathcal{T}$, an impossibility, since the hulls are totally ordered (cf. Lemma 2.3.3), unless E is 0 or I. It follows that A is a scalar multiple of I.

2.2. Some examples. The most familiar instances of maximal triangular algebras which are not finite dimensional arise from particular total orderings of an orthonormal basis for separable Hilbert space. They consist of all operators leaving invariant each of the subspaces generated by the basis vectors preceding a given one. These subspaces and those spanned by a basis vector and all basis vectors preceding it are the hulls—the diagonal is totally atomic (cf. Theorem 3.2.1). The operators leaving invariant the multiplication operators corresponding to the characteristic functions of intervals with left endpoint 0 on $L_2(0,1)$ relative to Lebesgue measure provide an example of a maximal triangular algebra with non-atomic diagonal. This same example relative to a measure with some atoms gives rise to a maximal triangular algebra with mixed diagonal. In each of these examples, two properties of the maximal triangular algebras are prominent: the core is equal to the diagonal and the hulls form a totally-ordered family (under the usual projection ordering). Examination of the finite-dimensional situation would lead us to suspect that these properties are valid for all maximal triangular algebras. In point of fact, however, the first does not hold in general (though it does for algebras with totally-atomic diagonals—cf. Theorem 3.2.1) while the second does (cf. Lemma 2.3.3). The theorem which follows provides us with the basis for specific examples of maximal triangular algebras whose core consists of the scalars (i.e. whose hulls are 0 and I).

We say that a unitary operator, U, acts ergodically on a von Neumann algebra, $\mathcal{A}$, when $U\mathcal{A}U^* = \mathcal{A}$ and there are no projections in $\mathcal{A}$ invariant under U other than 0 and I. In particular, no projections of $\mathcal{A}$ other than 0 and I commute with U; however, this is not equivalent to the ergodicity of U on $\mathcal{A}$. In fact, with $\{y_n\}_{n=0,\pm1,\cdots}$ an orthonormal basis, $F_n = [y_m : m \leq n]$, and $\mathcal{A}$ all bounded diagonal matrices; if we define U by: $Uy_n = y_{n-1}$; then U leaves each F_n invariant but commutes with no projection in $\mathcal{A}$, other than 0 and I.

THEOREM 2.2.1. *If the unitary operator, U, acts ergodically on the infinite-dimensional, maximal abelian algebra, $\mathcal{A}$, then the algebra, $\mathcal{S}$, generated by $\mathcal{A}$ and U is triangular, and the set of hulls in $\mathcal{S}$ is $\{0, I\}$; so that the same is true of each maximal triangular algebra containing $\mathcal{S}$.*

Proof. Since U acts ergodically on $\mathcal{A}$, we know that 0 and I are the only hulls, once we know that $\mathcal{S}$ is triangular. Note that each element, T, of $\mathcal{S}$ has the form $A_0 + A_1U + \cdots + A_nU^n$, with A_j in $\mathcal{A}$, since $U^nA = U^nAU^{-n}U^n = A'U^n$. Assume that T is self-adjoint, so that

$$A_0 + A_1U + \cdots + A_nU^n = A_0^* + U^{-1}A_1^* + \cdots + U^{-n}A_n^*.$$

Multiplying both sides by U^n, renormalizing, and transposing, we have $0 = A_0' + A_1'U + \cdots + A_nU^{2n}$, where A_n is as before. Assume that $A_0 + \cdots + A_nU^n = 0$ is an equation of minimal degree for U over $\mathcal{A}$. Let E be the range projection of A_n and F any projection in $\mathcal{A}$ contained in E. Then, if $G = U^nFU^{-n} \nleqq F$, we have $0 \neq G - GF (= M)$; whence $0 \neq U^{-n}MU^n \leqq F$. Writing N for $U^{-n}MU^n$, we have $NA_n \neq 0$, since N is a non-zero projection in the range of A_n; and $NA_0 + \cdots + NA_nU^n = 0$. Thus $NA_0 \neq 0$, for otherwise, U would satisfy an equation of degree lower than n over $\mathcal{A}$. But

$$\begin{aligned} 0 &= NA_0 + NA_1UN + \cdots + NA_nU^nN = NA_0 + \cdots + A_nNU^nNU^{-n}U^n \\ &= NA_0 + \cdots + A_nNMU^n = NA_0 + A_1'U + \cdots + A_{n-1}'U^{n-1} \end{aligned}$$

(recall that $FM = 0$ and $N \leqq F$). Since $NA_0 \neq 0$, this last equation contradicts the minimal property of n. Thus U^n leaves each projection in $\mathcal{A}$ contained in E invariant (i.e. in the present instance, $G \leqq F$).

The range projections of A_0 and A_n must be identical, for if they are not, since they commute, one contains a non-zero projection, G, in $\mathcal{A}$ orthogonal to the other, so that one of GA_0, GA_n, is 0 while the other is not. In either case, U would satisfy an equation of degree less than n. Thus, E is the range projection of both A_0 and A_n. From our equation for U, we obtain $A_0U^{-n} + A_1U^{-(n-1)} + \cdots + A_n = 0$; and this is an equation of lowest degree for U^{-1} over $\mathcal{A}$. In fact, if there is one of lower degree, by taking adjoints and renormalizing, we locate an equation of degree less than n satisfied by U. From the result of the preceding paragraph, therefore, U^{-n} $(= U^{n*})$ leaves each subprojection of E in $\mathcal{A}$ invariant. We conclude that U^n commutes with each such subprojection. Thus, with $F \leqq E$ and F in $\mathcal{A}$,

$$F + U^{-1}FU + \cdots + U^{-(n-1)}FU^{n-1}$$

commutes with U. By ergodicity of U on $\mathfrak{A}$, $F + \cdots + U^{-(n-1)}FU^{n-1}$ is a scalar multiple, $k_F I$, of I. Now each summand of

$$F + U^{-1}FU + \cdots + U^{-(n-1)}FU^{n-1}$$

is a projection, and these projections commute. Elementary spectral theory tells us that k_F is a positive integer ($F \neq 0$). If $E \neq F$, $k_F + k_{E-F} = k_E$, so that one of k_F, k_{E-F} does not exceed $k_E/2$. If F and $E - F$ are not minimal in $\mathfrak{A}$ and F, let us say, is such that $k_F \leqq k_E/2$, then we can choose F_1 in $\mathfrak{A}$, $F_1 < F$ such that $k_{F_1} \leqq k_E/4$. Thus, if E contains no minimal projections, we can locate a projection, F_m, in $\mathfrak{A}$, $F_m \leqq E$ with $k_{F_m} < k_E/2^m < 1$: a contradiction. Thus E contains a minimal projection, F, of $\mathfrak{A}$. It follows that U^jFU^{-j} are minimal in $\mathfrak{A}$, and $\mathfrak{A}$ being abelian, are (mutually) orthogonal or identical. Let m be the least integer such that $U^mFU^{-m} = F$, so that $m \leqq n$, F, $UFU^{-1} + \cdots + U^{m-1}FU^{-(m-1)}$ are orthogonal, and

$$F + UFU^{-1} + \cdots + U^{m-1}FU^{-(m-1)}$$

commutes with U. This sum is a scalar multiple of I, non-zero, and a projection. It is, therefore, I; so that $\mathfrak{A}$ is m-dimensional, contrary to hypothesis. It follows that U satisfies no polynomial equation over $\mathfrak{A}$ and $\mathfrak{S}$ is triangular. The maximal triangular algebras containing $\mathfrak{S}$ can't have a larger family of hulls than $\mathfrak{S}$ has; whence the hulls of such maximal triangular algebras are 0 and I.

This result and proof hold also in a factor.

Example 2.2.2. Let C be the unit circle with Lebesgue measure, $\mathfrak{H}$ be $L_2(C)$, $\mathfrak{A}$ be the multiplication algebra of $L_2(C)$, and U the unitary transformation of $\mathfrak{H}$ induced by an irrational rotation of C. It is well known that an irrational rotation of C is ergodic with respect to Lebesgue measure, from which we deduce that U acts ergodically on $\mathfrak{A}$. Of course, $\mathfrak{A}$ is infinite dimensional, whence from Theorem 2.2.1, $\mathfrak{A}$ and U generate a triangular algebra, $\mathfrak{S}$. Each maximal triangular algebra containing $\mathfrak{S}$ has 0 and I as its only hulls.

DEFINITION 2.2.3. *A triangular algebra whose only hulls are* 0 *and* I *will be said to be "irreducible."*

Irreducibility is equivalent to the core's consisting of scalars. It is clear, moreover, that the irreducible triangular algebras are those which act irreducibily on the underlying Hilbert space.

2.3. Some basic lemmas. In this section we develop a criterion which guarantees the membership of a given bounded operator in a particular maximal triangular algebra. With the aid of this result, we show that the hulls of maximal triangular algebras in factors are totally ordered (under the projection ordering).

Remark 2.3.1. If N is a projection invariant under the algebra $\mathcal{S}$ and M is a projection orthogonal to N, then $MSN = 0$ for each S in $\mathcal{S}$. In fact, $MSN = MNSN = 0SN = 0$.

LEMMA 2.3.2. *If $\mathfrak{A}$ is a self-adjoint operator algebra, $\mathcal{S}$ an operator algebra maximal with the property of having $\mathfrak{A}$ as its intersection with its adjoint, N and M orthogonal projections with N invariant under, and in $\mathcal{S}$, and B an operator such that $B = NBM$, then B lies in $\mathcal{S}$.*

Proof. Since $B^2 = 0$, each operator, T, in $\mathcal{S}_0$, the algebra generated by $\mathcal{S}$ and B has the form,

$$(*) \quad S + \sum BS \cdots S + \sum S \cdots SB + \sum BS \cdots SB + \sum SB \cdots S,$$

where the terms S which appear are not necessarily the same but all lie in $\mathcal{S}$. We shall show that $\mathcal{S}_0 \cap \mathcal{S}_0{}^* = \mathfrak{A}$, whence $\mathcal{S}_0 = \mathcal{S}$, by maximality, and B lies in $\mathcal{S}$. To this end, it will suffice, of course, to show that each self-adjoint operator T in $\mathcal{S}_0$ (so, in $\mathcal{S}_0 \cap \mathcal{S}_0{}^*$) lies in $\mathfrak{A}$, since $\mathcal{S}_0 \cap \mathcal{S}_0{}^*$ is a self-adjoint algebra and, therefore, generated by its self-adjoint operators.

We assume that T in $\mathcal{S}_0$ is self-adjoint and has the form described in $(*)$. By hypothesis, the range of B is contained in N, so that B leaves N invariant, and, thus, $\mathcal{S}_0$ leaves N invariant. From Remark 2.3.1, we conclude that $(I-N)TN = 0$, whence from the self-adjointness of T, $NT(I-N) = 0$. It follows that $T = NTN + (I-N)T(I-N)$. Now $(I-N)T(I-N) = (I-N)S(I-N)$, since $(I-N)B = (I-N)SB = (I-N)SNB = 0$, by hypothesis on B and invariance of N under S. With $(I-N)S(I-N)$ self-adjoint, we conclude that $(I-N)T(I-N)$ lies in $\mathfrak{A}$. It suffices, therefore, to show that NTN lies in $\mathfrak{A}$. But $NTN = NSN + \sum NBS \cdots BSN + \sum NS \cdots SBN + \sum NBS \cdots SBN + \sum NS \cdots BSN = NSN$, since $BN = BSN = 0$; and NSN lies in $\mathfrak{A}$, being a self-adjoint operator (NTN) in $\mathcal{S}$.

In the above lemma, we may assume that $\mathcal{S}$ has the maximal property with respect to some algebra, $\mathfrak{M}$, containing it, provided that the B in question lies in $\mathfrak{M}$. If $\mathfrak{M}$ is a factor, then with the notation of Lemma 2.3.2, we may state:

LEMMA 2.3.3. *If E and F are projections in $\mathcal{S}$ invariant under $\mathcal{S}$,*

then one of E, F contains the other (i.e. the invariant projections in $\mathcal{S}$ are totally ordered).

Proof. From invariance of E under $\mathcal{S}$, and, in particular, F, we have $EFE = FE$; whence, taking adjoints, $FE = EF$. Both $F - EF$ and $E - EF$ are non-zero unless one of E, F contains the other. If both are non-zero, there is a non-zero partial isometry, V, in $\mathcal{M}$ with initial space in $F - EF$ and final space in $E - EF$. Thus $V = EV(F - EF)$, with E invariant under $\mathcal{S}$ and orthogonal to $F - EF$. It follows, from the preceding lemma, that V lies in $\mathcal{S}$. However, V maps a part of F in $F - EF$ into $E - EF$, orthogonal to F, contradicting the invariance of F under $\mathcal{S}$. Thus, one of E, F contains the other.

Note that if $\mathcal{A}$ is generated by its invariant projections, the first statement of the foregoing proof shows that $\mathcal{A}$ is abelian.

LEMMA 2.3.4. *If $\mathcal{T}$ is a maximal triangular algebra in a factor $\mathcal{M}$ with diagonal $\mathcal{A}$ and core $\mathcal{B}$, then $h(G) - G$ is the hull immediately preceding $h(G)$ in $\mathcal{T}$ if G is a minimal projection in $\mathcal{B}$. If E is a hull in $\mathcal{T}$ which has a hull, F, immediately preceding it, then $E - F$ is a minimal projection in $\mathcal{B}$.*

Proof. If N is a hull in $\mathcal{T}$ not containing G, then $N \leqq h(G)$, from Lemma 2.3.3, and $GN = 0$, from minimality of G; so that $N \leqq h(G) - G$. The union, F, of all such hulls, N, is clearly a hull, $F \leqq h(G) - G$, and F is a hull immediately preceding $h(G)$. If we have proved the last statement of this lemma, then $h(G) - F$ is a minimal projection in $\mathcal{B}$ containing G, from which $F = h(G) - G$, and $h(G) - G$ is a hull. It remains to establish the last assertion of this lemma.

If M is a non-zero proper subprojection of $E - F$ in $\mathcal{B}$ and $N = E - F - M$, we can find a partial isometry, V, with initial space a non-zero subprojection of M and final space in N. If P is a hull of $\mathcal{T}$ containing E, then V and V^* leave P invariant. If P does not contain E, then $P \leqq F$; so that V and V^* annihilate P. Thus V commutes with each hull of $\mathcal{T}$ and hence with $\mathcal{B}$. However, V does not commute with M and M was chosen in $\mathcal{B}$. Thus $E - F$ is minimal in $\mathcal{B}$.

2.4. Other directions. The method by which we established the existence of maximal triangular algebras (a Zorn's Lemma construction) would yield, as well, the existence of an algebra, $\mathcal{T}$, maximal with respect to the property that $\mathcal{T}^* \cap \mathcal{T}$ is a given self-adjoint algebra, $\mathcal{A}$. More particularly,

we may choose $\mathfrak{A}$ to be abelian or to be a von Neumann algebra—the case where $\mathfrak{A}$ is maximal abelian is that of the maximal triangular algebras. (All this may be done in a given set of operators, e. g. a factor.) Without specifying $\mathfrak{A}$, we can construct an algebra, $\mathfrak{T}$, maximal with respect to the property that $\mathfrak{T}^* \cap \mathfrak{T}$ is abelian—and containing a given such algebra, $\mathfrak{T}_0$. In fact, take $\mathfrak{T}$ as the union of a maximal family of such algebras containing $\mathfrak{T}_0$ and totally ordered by inclusion. Clearly, $\mathfrak{T}$ is an algebra and $\mathfrak{T}^* \cap \mathfrak{T}$ is generated by its self-adjoint elements. If A_1 and A_2 are self-adjoint operators in $\mathfrak{T}$, then $A_1 \in \mathfrak{T}_1$, $A_2 \in \mathfrak{T}_2$, where $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are in the maximal family of which $\mathfrak{T}$ is the union. Say, $\mathfrak{T}_1 \subseteq \mathfrak{T}_2$, so that A_1 and A_2 lie in the abelian algebra, $\mathfrak{T}_2^* \cap \mathfrak{T}_2$. Thus $\mathfrak{T}^* \cap \mathfrak{T}$ is abelian. Of course $\mathfrak{T}_0^* \cap \mathfrak{T}_0 \subseteq \mathfrak{T}^* \cap \mathfrak{T}$, so that the abelian intersection with which we start may expand as we pass to the maximal algebra, $\mathfrak{T}$. Indeed, it would appear possible that all such maximal algebras are maximal triangular (i. e. $\mathfrak{T}^* \cap \mathfrak{T}$ is maximal abelian). We shall note that this need not be the case—even in finite dimensions. Before doing this, however, we wish to point out the importance of these considerations for certain critical questions in the theory of triangular operator algebras.

The question of whether or not a bounded operator on a separable Hilbert space has proper invariant subspaces may be strengthened and weakened in various ways. In a stronger form, one might ask not just for proper invariant subspaces, but for a "thick" family of such subspaces. A sense in which we can make this precise is to require that there be a resolution of the identity consisting of invariant subspaces, and more hopefully, a resolution with simple spectrum. Phrased in the language of our theory, we may ask:

Question 2.4.1. Is each bounded operator contained in some hyper-reducible maximal triangular algebra (core = diagonal—cf. Definition 3.0)?

This would provide a "triangular form" for bounded operators. We are inclined to feel that there is little hope that this question has an affirmative answer. It raises, in a natural way, the following question:

Question 2.4.2. Is each bounded operator contained in some maximal triangular algebra?

This is a much broader question, allowing, as it does, the possibility that the operator falls in an irreducible maximal triangular algebra. From this we would not conclude the existence of a single proper invariant subspace, although in the general sense of our theory, we would have the operator in "triangular form" and might gain knowledge about it from an analysis of the maximal triangular algebra in question. (The ostensibly weaker demand that

the operator lie in some triangular algebra is not really weaker, since each such algebra is contained in one which is maximal.) Suppose that B is a bounded operator and $\mathfrak{T}_0$ the (commutative) algebra generated by B and I (we may use any of the standard closures of $\mathfrak{T}_0$). Of course, $\mathfrak{T}_0{}^* \cap \mathfrak{T}_0$ is abelian, so that $\mathfrak{T}_0$ is contained in an algebra, $\mathfrak{T}$, maximal with respect to the property that $\mathfrak{T}^* \cap \mathfrak{T}$ is abelian. The example which follows shows that $\mathfrak{T}^* \cap \mathfrak{T}$ need not be maximal abelian.

Example 2.4.3. Let $\mathfrak{T}_1$ be the algebra of all 3×3 matrices, (a_{ij}), with $a_{31} = a_{32} = a_{13} = a_{23} = 0$, and let H be the positive square root of

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

For $\mathfrak{T}$ we choose $H\mathfrak{T}_1H^{-1}$. A computation shows that with A in $\mathfrak{T}_1$, HAH^{-1} is self-adjoint if and only if the matrix, (a_{ij}), for A, has $a_{ij} = 0$, $i \neq j$, and $a_{22} = a_{33}$. Thus the self-adjoint operators in $\mathfrak{T}$ are the image under an automorphism on 3×3 matrices of a 2-dimensional abelian set; and $\mathfrak{T}^* \cap \mathfrak{T}$ is not maximal abelian though abelian. It remains to show that $\mathfrak{T}$ is maximal with respect to the property of having an abelian intersection with its adjoint. We do this with a dimension argument. Suppose that $\mathcal{S}$ is a subalgebra of the 3×3 matrices which has dimension n. Let $\mathcal{S} \cup \mathcal{S}^*$ be the linear space spanned by $\mathcal{S}$ and $\mathcal{S}^*$. Then

$$2n = \dim(\mathcal{S}^* \cap \mathcal{S}) + \dim(\mathcal{S}^* \cup \mathcal{S}) \leqq \dim(\mathcal{S}^* \cap \mathcal{S}) + 9.$$

Since $\mathcal{S}^* \cap \mathcal{S}$ is self-adjoint, it has dimension 3 or less if it is abelian; so that $\mathcal{S}$ has dimension 6 or less. The algebra generated by $\mathfrak{T}_1$ and an operator not in it is easily seen to be at least 7 dimensional, whence the same is true for the automorph, $\mathfrak{T}$, of $\mathfrak{T}_1$. Thus, no algebra containing $\mathfrak{T}$ properly can have an abelian intersection with its adjoint.

Presumably, another computation would yield an example of an algebra with the maximal property having the scalars as intersection with its adjoint. With the aid of the example just constructed, we can produce an example in which the disparity betwen the dimension of the abelian intersection and that of the algebra is greater. In fact, let $\mathfrak{T}_0$ be the algebra of $m \times m$ matrices (corresponding to bounded operators, when $m = \infty$) whose entries are 3×3 matrices having each 3×3 entry below the diagonal the zero matrix and each 3×3 entry on the diagonal some matrix in $\mathfrak{T}$. Each self-adjoint operator in $\mathfrak{T}_0$ has non-zero entries only on the diagonal and these some self-adjoint element of $\mathfrak{T}$. Thus $\mathfrak{T}_0{}^* \cap \mathfrak{T}_0$ is abelian and $2m$-dimensional.

With m finite, maximal abelian algebras have dimension $3m$ and $\mathcal{T}_0$ has dimension $5m + \frac{9}{2}m(m-1)$. That $\mathcal{T}_0$ is maximal with respect to the property of having an abelian intersection with its adjoint is immediate from the corresponding fact for $\mathcal{T}$ and the fact that arbitrary entries appear above the diagonal.

A possible way of relating a given bounded operator to a maximal triangular algebra would be to replace the "part of the operator below the diagonal" by zero. Caution must be exercised with this process in infinite dimensions. Even in the classical case of the maximal triangular algebra arising from an orthonormal basis, $\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots$, the "super-diagonal" part of the operator may not be bounded. We may view this situation in terms of the group, $\mathfrak{I}$, of integers and $L_2(\mathfrak{I})$, relative to the disecrete measure on $\mathfrak{I}$ (Haar-Lebesgue measure). The Fourier-Plancherel transform establishes a unitary equivalence of $L_2(\mathfrak{I})$ with $L_2(C)$, where C is the circle group, which carries the maximal abelian algebra consisting of bounded L_2 convolution operators on $L_2(\mathfrak{I})$ onto the multiplication algebra of $L_2(C)$. If x_n is the function which is 1 at n and zero elsewhere on $\mathfrak{I}$, then the m, n-th entry of the matrix corresponding to convolution by f, relative to the basis $\{x_n\}$ for $L_2(\mathfrak{I})$, is $f(m-n)$, for $(f*x_n, x_m) = (f*x_n)(m) = f(m-n)$. If f_0 is 0 at positive integers and equal to f elsewhere on $\mathfrak{I}$, then the matrix for convolution by f_0 has as m, n-th entry (relative to $\{x_n\}$) $f_0(m-n)$, i.e. its matrix is the "super-diagonal part" of the matrix for f (when $n \geqq m$ the entry is $f(m-n)$, otherwise it is 0). However, while convolution by f may be a bounded operator, i.e. have Fourier-Plancherel transform a bounded measurable function on C, the transform of f_0 may be unbounded.

In general, then, we must not expect an operator, A, to have a decomposition, $A_1 + A_2$, with A_1 in $\mathcal{T}$ and A_2 in $\mathcal{T}^*$, where $\mathcal{T}$ is a given maximal triangular algebra. If such a decomposition exists, however, it is clearly unique up to an additive factor from the diagonal of $\mathcal{T}$.

In the chapter which follows, we shall give a reasonably detailed description of the most accessible class of maximal triangular algebras. This description will include an effective test of maximality. When we leave this class, no such test is known to us; and we may ask:

Question 2.4.4. Is there an easily applicable test for the maximality of a triangular algebra?

We have in mind some test such as that afforded by the Double Commutant Theorem in the theory of von Neumann algebras for the property of being strongly closed.

Theorem 2.2.1 and the example which follows it make use of Zorn's Lemma so that we do not have an explicit description of the operators in the irreducible maximal triangular algebra which results.

Question 2.4.5. Is there an explicit construction of an irreducible maximal triangular algebra?

Questions 2.4.4 and 2.4.5 are admittedly vague though, nonetheless, important for this theory. The following very definite question is perhaps the most provoking sample from a long list of questions one could ask about the irreducible triangular algebras.

Question 2.4.6. Are there two (or more) irreducible maximal triangular algebras on separable Hilbert space which are not algebraically isomorphic?

Chapter III. Hyperreducible Algebras.

DEFINITION 3.0. *A triangular algebra whose hulls generate the diagonal is said to be hyperreducible.*

Note that each hull of a triangular algebra is a reducing subspace—the hyperreducible case is the one with the greatest possible reduction. Hyperreducible algebras are those for which the core is equal to the diagonal. We have noted (cf. Remark 2.1.5) that the core plays the rôle of the center, so that the hyperreducible algebras in the theory of triangular operator algebras would correspond to the abelian algebras of the self-adjoint theory. Through this analogy, we would expect the hyperreducible algebras to be the most tractable of the triangular algebras, and this is the case—though their theory is not nearly as complete at this time as the abelian self-adjoint theory.

3.1. The general structure. The following result gives, in very broad terms, the general structure of the maximal hyperreducible algebras.

THEOREM 3.1.1. *If $\{E_\alpha\}$ is a totally-ordered family of projections which generates the maximal abelian algebra, $\mathfrak{A}$; then $\mathfrak{T}$, the set of all bounded operators which leave each E_α invariant is a maximal triangular algebra, with core and diagonal $\mathfrak{A}$. If $\{E_\alpha\}$ is closed under unions and intersections then it is the set of hulls of $\mathfrak{T}$. Each hyperreducible maximal triangular algebra arises in this way.*

Proof. That $\mathfrak{T}$ is an algebra is clear. If A is a self-adjoint operator in $\mathfrak{T}$, then, since $AE_\alpha = E_\alpha A E_\alpha = (E_\alpha A E_\alpha)^* = E_\alpha A$, we conclude that A

commutes with and hence lies in $\mathcal{A}$, so that $\mathcal{T}$ is triangular with diagonal (and core, by hypothesis) equal to $\mathcal{A}$.

Choose a maximal triangular algebra, $\mathcal{T}_0$, containing $\mathcal{T}$, and let B be an operator in $\mathcal{T}_0$. For each operator, T, $E_\alpha T(I-E_\alpha)$ lies in $\mathcal{T}$. Indeed, with $E_\beta \leqq E_\alpha$,

$$0 = E_\alpha T(I-E_\alpha)E_\beta = E_\beta E_\alpha T(I-E_\alpha)E_\beta,$$

while, with $E_\beta \geqq E_\alpha$,

$$E_\alpha T(I-E_\alpha)E_\beta = E_\beta E_\alpha T(I-E_\alpha)E_\beta.$$

In particular, $E_\alpha B^*(I-E_\alpha)$ lies in $\mathcal{T}$ (hence in $\mathcal{T}_0$). It follows that the self-adjoint operator,

$$E_\alpha B^*(I-E_\alpha) + (I-E_\alpha)BE_\alpha,$$

lies in $\mathcal{T}_0$ and, so, in $\mathcal{A}$. Commutativity with E_α then gives

$$E_\alpha B^*(I-E_\alpha) = (I-E_\alpha)BE_\alpha,$$

from which, by multiplying both sides by $(I-E_\alpha)$, we conclude that $(I-E_\alpha)BE_\alpha = 0$. Thus B leaves each E_α invariant, B lies in $\mathcal{T}$, $\mathcal{T} = \mathcal{T}_0$, and $\mathcal{T}$ is a maximal triangular algebra.

On the other hand, if $\mathcal{T}$ is maximal triangular and hyperreducible with hulls $\{E_\alpha\}$ and diagonal $\mathcal{A}$, then $\{E_\alpha\}$ generates the maximal abelian algebra, $\mathcal{A}$, and is totally ordered. Thus, $\mathcal{T}_0$, the set of operators leaving each E_α invariant, is maximal triangular and contains $\mathcal{T}$. By maximality of $\mathcal{T}$, we have, $\mathcal{T} = \mathcal{T}_0$.

If $\{E_\alpha\}$ is closed under union and intersection of its members, and E is a hull for $\mathcal{T}$, then E_0 the union of, and E_1 the intersection of all E_α contained in and containing E, respectively, lie in $\{E_\alpha\}$. If $E_0 < E < E_1$ then, as in Lemma 2.3.4, a partial isometry with initial space in $E - E_0$ and range in $E_1 - E$ commutes with each E_α, hence with $\mathcal{A}$, but does not leave $E - E_0$ invariant. Thus E is one of E_0, E_1, and E lies in $\{E_\alpha\}$. The hulls of $\mathcal{T}$ are precisely the E_α, in this case.

3.2. Triangular algebras with totally-atomic diagonals. Throughout this section, we shall be discussing the maximal triangular algebra, $\mathcal{T}$, over a maximal abelian algebra $\mathcal{A}$ which is generated by its minimal projections. In this case, the structure of $\mathcal{T}$ can be completely described. The main result is contained in:

THEOREM 3.2.1. *If $\mathcal{T}$ is a maximal triangular algebra with diagonal*

$\mathcal{A}$ which is generated by its minimal projections, then $\mathcal{T}$ is hyperreducible. The total ordering of the hulls induces a total ordering on the minimal projections, $\{E_\alpha\}$, of $\mathcal{A}$ by means of the mapping from projections to their hulls (which is one-one on the minimal projections); two such triangular algebras are unitarily equivalent if and only if their sets of minimal projections are order isomorphic. Corresponding to each total-ordering type there is a maximal triangular algebra with a totally-atomic diagonal whose set of minimal projections has this order type.

Proof. Since $h(E_\alpha) = [\mathcal{T}E_\alpha]$, $E_\beta h(E_\alpha) \neq 0$ if and only if $[E_\beta \mathcal{T} E_\alpha] \neq 0$, i. e. $E_\beta T E_\alpha \neq 0$, for some T in $\mathcal{T}$. If, in addition, $E_\alpha h(E_\beta) \neq 0$, then $E_\alpha T' E_\beta \neq 0$, for some T' in $\mathcal{T}$. Now $E_\alpha T^* E_\beta$ is a scalar multiple of $E_\alpha T' E_\beta$, since E_α and E_β are one dimensional, so that $E_\alpha T^* E_\beta + E_\beta T E_\alpha$ lies in $\mathcal{A}$, and commutes with E_α, E_β,—contradicting $E_\beta T E_\alpha \neq 0$. Thus $E_\alpha h(E_\beta) = 0$, and $h(E_\beta) \leqq h(E_\alpha) - E_\alpha$. Since $h(E_\alpha) - E_\alpha = \sum E_\beta$, $h(E_\alpha) - E_\alpha = \vee\, h(E_\beta)$. Hence $h(E_\alpha) - E_\alpha$ is a hull, so that $E_\alpha (= h(E_\alpha) - [h(E_\alpha) - E_\alpha])$ lies in the core of $\mathcal{T}$. Since $\mathcal{A}$ is generated by $\{E_\alpha\}$, $\mathcal{T}$ is hyperreducible. If $h(E_\alpha) = h(E_\beta)$, then $h(E_\alpha) - E_\alpha = h(E_\beta) - E_\beta$; and $E_\alpha = E_\beta$. In fact, from Lemma 2.3.4, $h(E_\alpha) - E_\alpha$ and $h(E_\beta) - E_\beta$ are the hulls immediately preceding $h(E_\alpha)$ and $h(E_\beta)$, respectively. Thus the total ordering of the hulls induces a total ordering, $\ll$, of $\{E_\alpha\}$ by means of the one-one mapping, $h(E_\alpha) \to E_\alpha$. Since each projection in $\mathcal{A}$ is the sum of the minimal projections it contains, and $E_\beta \ll E_\alpha$ (i.e. $h(E_\beta) \leqq h(E_\alpha)$) if and only if $E_\beta \leqq h(E_\alpha)$; we have, $h(E_\alpha) = \sum_{E_\beta \ll E_\alpha} E_\beta$.

If $\mathcal{T}_1$ ond $\mathcal{T}_2$ are maximal triangular algebras with diagonals $\mathcal{A}_1$ and $\mathcal{A}_2$ which are generated by their sets $\{E_\alpha\}$, $\{F_\alpha\}$ of minimal projections, acting on the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, and the mapping $E_\alpha \to F_\alpha$ is an order isomorphism of $\{E_\alpha\}$ onto $\{F_\alpha\}$ relative to the $\ll$ ordering on these sets; then the unitary transformation, U, defined by mapping a unit vector in the range of E_α onto a unit vector in the range of F_α is such that $U\mathcal{A}_1U^{-1} = \mathcal{A}_2$ and $UE_\alpha U^{-1} = F_\alpha$. We shall show that $U\mathcal{T}_1U^{-1} = \mathcal{T}_2$. Indeed, from the description of $\mathcal{T}_1$ and $\mathcal{T}_2$ as the algebras of all operators on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, which leave the hulls in $\mathcal{T}_1$ and $\mathcal{T}_2$ invariant, it suffices to show that UEU^{-1} is a hull in $\mathcal{T}_2$ if E is a hull in $\mathcal{T}_1$ (by symmetry, then, F is a hull in $\mathcal{T}_1$ if UFU^{-1} is a hull in $\mathcal{T}_2$). Now $Uh(E_\alpha)U^{-1} = h(F_\alpha)$, since $h(E_\alpha) = \sum_{E_\beta \ll E_\alpha} E_\beta$, $h(F_\alpha) = \sum_{F_\beta \ll F_\alpha} F_\beta$, and $E_\beta \ll E_\alpha$ if and only if $F_\beta = UE_\beta U^{-1} \ll F_\alpha = UE_\alpha U^{-1}$. If E is a hull in $\mathcal{T}_1$, then $E = \bigvee_{E_\alpha \leqq E} h(E_\alpha)$, whence $UEU^{-1} = \bigvee_{UE_\alpha U^{-1} \leqq UEU^{-1}} Uh(E_\alpha)U^{-1} = \bigvee_{F_\alpha \leqq UEU^{-1}} h(F_\alpha)$; and UEU^{-1} is a hull in $\mathcal{T}_2$.

Let $\{\alpha\}$ be a set having a given total-order type, $\mathcal{H}$ a Hilbert space having dimension the cardinality of $\{\alpha\}$, and let $\{E_\alpha\}$ be a maximal orthogonal family of one-dimensional projections on $\mathcal{H}$ indexed by $\{\alpha\}$ and totally ordered by the relation, $<<$, induced by this indexing. If $\mathcal{A}$ is the maximal abelian algebra generated by $\{E_\alpha\}$ and E is a projection in $\mathcal{A}$, define $h(E)$ to be $\bigvee\{E_\beta\colon E_\beta << E_\alpha$ for some $E_\alpha \leqq E\}$. Note that the set of β's involved in the defining sum for $h(E)$ is an initial segment of $\{\alpha\}$. Since initial segments of a totally-ordered set form a totally-ordered set, the projections in $\{h(E)\colon E$ a projection in $\mathcal{A}\}$ form a totally-ordered set. The definition implies that $h(E_\alpha) = \sum_{\beta<<\alpha} E_\beta$; whence $h(E_\alpha) - E_\alpha = h(h(E_\alpha) - E_\alpha)$. Thus, each E_α lies in the algebra generated by $\{h(E)\}$, and this algebra is, therefore, $\mathcal{A}$. From Theorem 3.1.1, the set of operators, $\mathcal{T}$, leaving each $h(E)$ invariant is maximal, hyperreducible. Moreover, the equality, $h(E_\alpha) = \sum_{\beta<<\alpha} E_\beta$, establishes the fact that h is an order isomorphism of $\{E_\alpha\}$ with $\{h(E_\alpha)\}$; and the proof is complete.

Corollary 3.2.2. *If $\mathcal{T}_1$ and $\mathcal{T}_2$ are maximal triangular algebras with totally-atomic diagonals and φ is an order isomorphism between their sets of hulls, then φ can be implemented by a unitary transformation which carries $\mathcal{T}_1$ onto $\mathcal{T}_2$.*

Proof. By virtue of Theorem 3.2.1, it will suffice to show that φ induces an order isomorphism between the sets of minimal projections. We recall that the mapping from minimal projections to their hulls is an order isomorphism between the minimal projections and their hulls. It remains to note that φ carries the hull, $h(G)$, of a minimal projection, G, in $\mathcal{T}_1$ onto such a hull in $\mathcal{T}_2$. Now $\varphi(h(G) - G)$ is a hull in $\mathcal{T}_2$ immediately preceding $\varphi(h(G))$; whence $\varphi(h(G))$ is the hull of a minimal projection $\mathcal{T}_2$ (cf. Lemma 2.3.4).

Theorem 3.2.3. *If $\mathcal{T}_1$ and $\mathcal{T}_2$ are maximal triangular algebras with totally-atomic diagonals, $\mathcal{A}_1$ and $\mathcal{A}_2$, acting on Hilbert spaces, $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, and φ is an isomorphism of $\mathcal{T}_1$ onto $\mathcal{T}_2$ carrying $\mathcal{A}_1$ onto $\mathcal{A}_2$, then φ is implemented by a bicontinuous linear isomorphism of $\mathcal{H}_1$ onto $\mathcal{H}_2$. In particular, if $\mathcal{T}_1 = \mathcal{T}_2$ and φ is the identity transform on $\mathcal{A}_1$, then the implementing transformation lies in $\mathcal{A}_1$.*

Proof. Since $\varphi(\mathcal{A}_1) = \mathcal{A}_2$, φ carries the set of minimal projections, $\{E_\alpha\}$, of $\mathcal{A}_1$ onto that of $\mathcal{A}_2$. Moreover, since a projection, E, is invariant under T if and only if $TE = ETE$, φ preserves hulls. Now $E_\alpha << E_{\alpha'}$ if and only if

$E_\alpha \leqq h(E_{\alpha'})$; whence $\varphi(E_\alpha) \leqq \varphi(h(E_{\alpha'})) = h(\varphi(E_{\alpha'}))$, and $\varphi(E_\alpha) \ll \varphi(E_{\alpha'})$. Applying these considerations to φ^{-1}, we conclude that φ induces an order isomorphism of the minimal projections in $\mathfrak{T}_1$ onto those of $\mathfrak{T}_2$; so that there is a unitary transformation of $\mathcal{H}_1$ onto $\mathcal{H}_2$ carrying $\mathfrak{T}_1$ onto $\mathfrak{T}_2$ and implementing φ on $\{E_\alpha\}$ (hence on $\mathcal{A}_1$). Composing the inverse of this unitarily induced mapping with φ, we see that it suffices to consider the case where φ is an automorphism of $\mathfrak{T}_1$ which is the identity transform on $\mathcal{A}_1$. Since $E_{\alpha'}\mathfrak{T}_1 E_\alpha$ is one-dimensional if $E_{\alpha'} \ll E_\alpha$,

$$\varphi(E_{\alpha'}TE_\alpha) = E_{\alpha'}\varphi(T)E_\alpha = a_{\alpha'\alpha}E_{\alpha'}TE_\alpha, \text{ where } a_{\alpha'\alpha}$$

is some non-zero scalar (independent of T). If $E_{\alpha''} \ll E_{\alpha'}$, then

$$\begin{aligned} a_{\alpha''\alpha}E_{\alpha''}T'E_{\alpha'}TE_\alpha &= \varphi(E_{\alpha''}T'E_{\alpha'}TE_\alpha) \\ &= \varphi(E_{\alpha''}T'E_{\alpha'})\varphi(E_{\alpha'}TE_\alpha) = a_{\alpha''\alpha'}a_{\alpha'\alpha}E_{\alpha''}T'E_{\alpha'}TE_\alpha; \end{aligned}$$

whence $a_{\alpha''\alpha'}a_{\alpha'\alpha} = a_{\alpha''\alpha}$. Clearly $a_{\alpha\alpha} = 1$, for all α. Fix α'', and define b_α to be $a_{\alpha''\alpha}^{-1}$ (or $a_{\alpha\alpha''}$ if $\alpha \ll \alpha''$). Let $B = \sum_\alpha b_\alpha E_\alpha$. We assert that B is an invertible operator in $\mathcal{A}_1$. To establish this, we need show only that $\{|a_{\alpha'\alpha}|\}$ is bounded above, for applying this result to φ^{-1}, we conclude that $\{|a_{\alpha'\alpha}|^{-1}\}$ is bounded above; whence $\{|b_\alpha|\}$ and $\{|b_\alpha^{-1}|\}$ are bounded above. If $\{|a_{\alpha'\alpha}|\}$ is not bounded above, then for each positive integer, n, there exist α_n and α_n', with $\alpha_n' \ll \alpha_n$ such that $|a_{\alpha_n'\alpha_n}| \geqq n^3$. Now $\sum_n (1/n^2)T_n$ converges uniformly to an operator, T, in $\mathfrak{T}_1$, where T_n is a partial isometry (in $\mathfrak{T}_1$) with initial space E_{α_n} and range $E_{\alpha_n'}$. But

$$\| \varphi(T) \| \geqq \| E_{\alpha_n'}\varphi(T)E_{\alpha_n} \| = \| (a_{\alpha_n'\alpha_n}/n^2)T_n \| \geqq n,$$

whence $\varphi(T)$ is not bounded, a contradiction. Thus $\{|a_{\alpha'\alpha}|\}$ is bounded. With T in $\mathfrak{T}_1$,

$$\begin{aligned} E_{\alpha'}(BTB^{-1})E_\alpha &= B(E_{\alpha'}TE_\alpha)B^{-1} = b_{\alpha'}b_\alpha^{-1}E_{\alpha'}TE_\alpha \\ &= a_{\alpha''\alpha'}^{-1}a_{\alpha''\alpha}E_{\alpha'}TE_\alpha = a_{\alpha'\alpha}E_{\alpha'}TE_\alpha = \varphi(E_{\alpha'}TE_\alpha) = E_{\alpha'}\varphi(T)E_\alpha. \end{aligned}$$

It follows that $BTB^{-1} = \varphi(T)$, and the proof is complete.

An orthonormal basis for a Hilbert space determines and is determined by the maximal abelian algebra of bounded operators with this basis as eigenvectors (the diagonal matrices relative to this basis). It is natural and customary therefore, to think of the arbitrary maximal abelian algebra as a generalized basis. In the same way, an ordered basis for the Hilbert space corresponds to the maximal triangular, hyperreducible algebra with diagonal

the maximal abelian algebra of this basis and atoms ordered by the basis. For this reason, we may think of the general maximal triangular, hyperreducible algebra as a generalized ordered basis, and we shall often refer to such an algebra as an "ordered basis." When the diagonal is totally atomic, we shall speak of a "discrete, ordered basis"; and when the atoms are ordered as the integers (positive, negative, or positive and negative), we shall speak of an "integer-ordered basis." The atoms of an integer-ordered basis correspond to those infinite, totally-ordered sets between each pair of elements of which, there are a finite number of elements.

3.3. Non-atomic hyperreducible algebras. In this section, we consider hyperreducible, maximal triangular algebras (on separable spaces) whose diagonals are non-atomic. We shall see that, unlike the totally-atomic case, all such algebras are unitarily equivalent and *a fortiori* isomorphic (cf. Theorem 3.3.1), but that not each order isomorphism between the hulls can be implemented by a unitary transformation.

THEOREM 3.3.1. *If $\mathcal{T}$ is a hyperreducible, maximal triangular algebra with non-atomic diagonal, $\mathcal{A}$, acting on a separable Hilbert space, then $\mathcal{T}$ is unitarily equivalent to $\mathcal{T}_0$, the algebra of all bounded operators on $L_2(0,1)$ (Lebesgue measure) leaving each F_λ invariant, where F_λ is the projection due to multiplication by the characteristic function, χ_λ, of $[0,\lambda]$.*

Proof. Since $\mathcal{A}$ is abelian on a separable space, there is a unit vector, x, which is separating for $\mathcal{A}$. The hulls being totally ordered and x separating, ω_x takes distinct values on distinct hulls so that we can index each hull, E, with $\omega_x(E)$. Let $\{E_\lambda\}$ be these hulls so indexed. Note that $E_0 = 0$, $E_1 = I$, and for each μ in $[0,1]$, there is an E_μ with $\bigwedge_{\lambda>\mu} E_\lambda = E_\mu = \bigvee_{\nu<\mu} E_\nu$. In fact, if this were not so, $\bigvee_{\nu<\mu} E_\nu$ would be the hull immediately preceding $\bigwedge_{\lambda>\mu} E_\lambda$ which would, according to Lemma 2.3.4, be the hull of a minimal projection. Thus $\{E_\lambda\}$ is a resolution of the identity, say $A = \int \lambda \, dE_\lambda$. Let $\mathcal{A}_0$ be the multiplication algebra of $L_2(0,1)$ (Lebesgue measure), and let $\varphi(f(A))$ be T_f, the operator due to multiplication by f, a bounded measurable function on $[0,1]$. Because of the properties of $\{E_\lambda\}$, φ is an isomorphism of $\mathcal{A}$ with $\mathcal{A}_0$. The mapping $f(A)x \to f$ is isometric; for

$$\begin{aligned}(f(A)x, g(A)x) &= \int \bar{g}(\lambda) f(\lambda) \, d(E_\lambda x, x) = \int \bar{g}(\lambda) f(\lambda) \, d\lambda \\ &= (f, g),\end{aligned}$$

and therefore has a unitary extension, U, which is easily seen to implement φ.

Now $X_\lambda(A) = E_\lambda$, whence $\varphi(E_\lambda) = T_{X_\lambda} = F_\lambda$. From Theorem 3.1.1, $\mathcal{T}$ is describable as the algebra of all bounded operators which leave each E_λ invariant, so that U carries $\mathcal{T}$ onto $\mathcal{T}_0$.

If f is an order isomorphism of $[0,1]$ onto $[0,1]$ the mapping $E_\lambda \to E_{f(\lambda)}$ will be an order isomorphism of the hulls of $\mathcal{T}$ onto itself (using the notation of the preceding theorem). We may ask ourselves if this mapping can be implemented by a unitary transformation of $\mathcal{T}$ onto $\mathcal{T}$. The corresponding question in the totally-atomic case had an affirmative answer. In the present case, the answer is in the negative. An indication of why this is so will serve as a good introduction to the methods of the following section. Such a unitary transformation must carry $\int \lambda \, dE_\lambda$ onto $\int \lambda \, dE_{f(\lambda)}$ $(= \int f^{-1}(\mu) \, dE_\mu)$. We note that, under the hypothesis, f will be a homeomorphism of $[0,1]$ onto itself (the image of an interval is an interval). Thus, the C^*-algebras generated by both $\int \lambda \, dE_\lambda$ and $\int f^{-1}(\lambda) \, dE_\lambda$ correspond to the algebra of multiplications by continuous functions on $[0,1]$, both have simple spectrum $[0,1]$, and the spectral null sets in both cases are given by the constant function, 1. In the first case, these are the null sets of the integration process, $g \to \int g(\lambda) \, d\lambda$, and in the second, those of $g \to \int g(f^{-1}(\lambda)) \, d\lambda$. The first, then, are the Borel subsets of $[0,1]$ of Lebesgue measure 0, and the second are the images under f of these. Our task, then, is to construct an f which does not preserve the Borel sets of measure 0. Such functions are described in the literature (cf., [3, p. 83]). For example, let f be defined by $f(\lambda) = (\lambda + g(\lambda))/2$, where g is the Cantor function.

3.4. The general diagonal. We consider, now, the case where the diagonal is not assumed pure in the sense of total atomicity or non-atomicity. Experience with the self-adjoint theory conditions us to expect that the general case is a simple matter of separating the diagonal into its totally-atomic and non-atomic parts. This is not so in the present theory, as will be evident from the results of this section. The order, which is a primary constituent of this investigation, places the atoms throughout the continuous portion of the diagonal in a manner which does not permit a separation consistent with this theory.

DEFINITION 3.4.1. *The "hull class," $h(\mathcal{T})$, of an ordered basis, $\mathcal{T}$, on a separable space with hulls, $\{E_\alpha\}$, and diagonal, $\mathcal{A}$, is the class $\{\{(E_\alpha x, x)\}$: x a separating vector for $\mathcal{A}\}$ of subsets of $[0,1]$.*

Remark 3.4.2. The mapping η, taking E_α onto $(E_\alpha x, x)$ is an order isomorphism since $\{E_\alpha\}$ is totally ordered and x is a separating vector for $\mathcal{A}$.

The greatest lower bound, E_{α_0}, of a subset of $\{E_\alpha\}$ is its intersection and a strong limit point of it; whence $\eta(E_{\alpha_0})$ is a lower bound and a limit point of its image under η. Thus $\eta(E_{\alpha_0})$ is the greatest lower bound of this image. It follows that the greatest lower bound (and likewise the least upper bound) of each subset of $\{(E_\alpha x, x)\}$ lies in it, whence $\{(E_\alpha x, x)\}$ is closed. Each member of the hull class is a closed subset of $[0,1]$ containing 0 and 1.

Remark 3.4.3. With X in $h(\mathcal{T})$ and X' the complement of X in $[0,1]$, X' is open and thus, the sum of a countable family, $I_1, I_2, \cdots$, of disjoint open intervals. The left and right hand endpoints, l_k and r_k, respectively, of I_k lie in X (we refer to the set of these endpoints as "the edge of X," to the l_k as "the left edge," and to the r_k as "the right edge"); and l_k corresponds to the hull immediately preceding the hull which corresponds to r_k. Thus r_k corresponds to the hull of a minimal projection in $\mathcal{A}$ (cf. Lemma 2.3.4). Conversely, if G is a minimal projection in $\mathcal{A}$, then $h(G) - G$ is the hull immediately preceding $h(G)$, so that the (non-empty) interval $(\eta(h(G) - G), \eta(h(G)))$ is some I_k. There is a one-one correspondence effected by η between the hulls of minimal projections in $\mathcal{A}$ and the right edge points of X.

THEOREM 3.4.4. *Two separable ordered bases, $\mathcal{T}_1$ and $\mathcal{T}_2$, are unitarily equivalent if $h(\mathcal{T}_1) \cap h(\mathcal{T}_2) \neq \phi$ and only if $h(\mathcal{T}_1) = h(\mathcal{T}_2)$.*

Proof. A unitary equivalence between $\mathcal{T}_1$ and $\mathcal{T}_2$ preserves diagonals, hulls, and the separating vectors for the diagonals; whence $h(\mathcal{T}_1) = h(\mathcal{T}_2)$.

Suppose, now, that x and y are separating vectors for the diagonals of $\mathcal{T}_1$ and $\mathcal{T}_2$ which give rise to the same set, X, in both $h(\mathcal{T}_1)$ and $h(\mathcal{T}_2)$. Let us index the hulls of $\mathcal{T}_1$ and $\mathcal{T}_2$ by points of X in such a way that $(E_\lambda x, x) = \lambda = (F_\lambda y, y)$, for each of the hulls E_λ and F_λ in $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. The sets $\mathcal{A}_1$ and $\mathcal{A}_2$ of linear combinations of $\{E_\lambda\}$ and $\{F_\lambda\}$, respectively, are self-adjoint algebras which are weakly dense in the diagonals of $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively, so that $[\mathcal{A}_1 x] = \mathcal{H}_1$ and $[\mathcal{A}_2 y] = \mathcal{H}_2$ (with $\mathcal{H}_1$ and $\mathcal{H}_2$ the Hilbert spaces upon which $\mathcal{T}_1$ and $\mathcal{T}_2$ act). The mapping taking $(\sum_{i=1}^n \alpha_i E_{\lambda_i})x$ onto $(\sum_{i=1}^n \alpha_i E_{\lambda_i})y$ is isometric, for

$$\|(\sum_{i=1}^n \alpha_i E_{\lambda_i})x\|^2 = (\sum_{i,j=1}^n \alpha_i \bar{\alpha}_j E_{\lambda_i} E_{\lambda_j} x, x)$$
$$= (\sum_{i,j=1}^n \alpha_i \bar{\alpha}_j F_{\lambda_i} F_{\lambda_j} y, y) = \|\sum_{i=1}^n \alpha_i F_{\lambda_i} y\|^2.$$

Thus, this mapping has a unitary extension, U. Note that,

$$UE_\lambda U^{-1}((\sum \alpha_i F_{\lambda_i})y) = UE_\lambda(\sum \alpha_i E_{\lambda_i})x = U(\sum \alpha_i E_\lambda E_{\lambda_i})x$$
$$= (\sum \alpha_i F_\lambda F_{\lambda_i})y = F_\lambda((\sum \alpha_i F_{\lambda_i})y);$$

whence $UE_\lambda U^{-1} = F_\lambda$, and U effects a unitary equivalence between $\mathfrak{T}_1$ and $\mathfrak{T}_2$ (cf. Theorem 3.1.1).

The concept of an orientation-preserving homeomorphism which carries the null sets of Lebesgue measure onto these null sets of the image will play an important role in our work. We shall call such a mapping "a Lebesgue order isomorphism."

THEOREM 3.4.5. *If X is a closed subset of $[0,1]$ containing 0 and 1, X' its complement in $[0,1]$ with connected components $I_1, I_2, \cdots, \mu$ is the Borel measure on $[0,1]$ defined by $\mu(S) = m(S \cap X) + \sum_{r_k \in S} m(I_k)$, where r_k is the right endpoint of I_k and m is Lebesgue measure on $[0,1]$, $\mathfrak{A}$ is the multiplication algebra of $L_2([0,1],\mu)$, $\mathfrak{T}$ is the algebra of all bounded operators on $L_2([0,1],\mu)$ leaving each E_λ and $E_{\lambda\cdot}$ invariant, where E_λ and $E_{\lambda\cdot}$ are the projections due to multiplication by the characteristic functions of the half-open interval, $[0,\lambda)$, and the closed interval, $[0,\lambda]$, respectively, then $\mathfrak{T}$ is an ordered basis with diagonal $\mathfrak{A}$ and hulls $\{E_\lambda, E_{\lambda\cdot}\}$, and $X \in h(\mathfrak{T})$.*

Proof. Each operator due to multiplication by the characteristic function of a closed or open interval, and hence, a Borel set, lies in the von Neumann algebra generated by $\{E_\lambda, E_{\lambda\cdot}\}$; whence this algebra is $\mathfrak{A}$. Of course, $\{E_\lambda, E_{\lambda\cdot}\}$ is a totally-ordered family of projections, and from Theorem 3.1.1, $\mathfrak{T}$ is an ordered basis with diagonal $\mathfrak{A}$. Let $\mathfrak{F}$ be a subset of $\{E_\lambda, E_{\lambda\cdot}\}$, E its intersection, and γ the right endpoint of the interval which is the intersection of the intervals corresponding to the projections of $\mathfrak{F}$. Clearly $E_\gamma \leqq E$. Now E corresponds to multiplication by the characteristic function of some μ-measurable subset, S, of $[0,1]$. If $\lambda > \gamma$, there is a λ', with $E_{\lambda'}$ or $E_{\lambda'\cdot}$ in $\mathfrak{F}$, such that $\gamma \leqq \lambda' < \lambda$; whence $E \leqq E_{\lambda'\cdot} \leqq E_\lambda$. Thus, $\mu(\{\lambda' : \lambda' \in S, \lambda' > \lambda\}) = 0$, and since this holds for each $\lambda > \gamma$, $\mu(\{\lambda : \lambda \in S, \lambda > \gamma\}) = 0$. Hence $E \leqq E_{\gamma\cdot}$, and $E = E_\gamma$ or $E_{\gamma\cdot}$. Similarly, the union of the projections in $\mathfrak{F}$ lies in $\{E_\lambda, E_{\lambda\cdot}\}$, and from Theorem 3.1.1, $\{E_\lambda, E_{\lambda\cdot}\}$ is the set of hulls of $\mathfrak{T}$.

Observe that each r_k lies in X, so that if $S \cap X = \phi$ then $\mu(S) = 0$. Note also that $\mu([0,\lambda_0]) = \lambda_0$, when $\lambda_0 \in X$, for

$$\mu([0,\lambda_0]) = m([0,\lambda_0] \cap X) + \sum_{r_k \leqq \lambda_0} m(I_k)$$
$$= m([0,\lambda_0] \cap X) + m([0,\lambda_0] - [0,\lambda_0] \cap X) = \lambda_0.$$

Moreover, $\mu([0,\lambda_0))=\lambda_0$ or l_k if $\lambda_0\notin\{r_k\}$ or $\lambda_0=r_k$, respectively. Now, if $\lambda\notin X$, $(E_\lambda x,x)=\mu([0,\lambda))=\mu([0,\lambda])=(E_{\lambda'}x,x)$, where x is the constant function 1 on $[0,1]$. If λ_0 is the least upper bound of $[0,\lambda]\cap X$,

$$\mu([0,\lambda))=\mu([0,\lambda_0])+\mu((\lambda_0,\lambda))=\mu([0,\lambda_0])=\lambda_0\in X.$$

Thus X is the member of the hull class of $\mathfrak{F}$ corresponding to the separating vector x.

The ordered bases arising from the constructions of the foregoing theorem contain representatives from each unitary equivalence class of ordered bases, since each ordered basis is unitarily equivalent to the ones constructed on each of the sets in its hull class. We must still say, however, when two ordered bases arising from the construction of Theorem 3.4.5 are unitarily equivalent (and which closed sets appear in a given hull class). Theorem 3.4.8 answers these questions. Several remarks will be of help.

Remark 3.4.6. If X and Y are closed subsets of $[0,1]$ containing 0 and 1, and f is an order isomorphism of X onto Y then f has an order-isomorphic extension mapping $[0,1]$ onto itself. In particular, the extension is a homeomorphism, and f is continuous on X. In fact, the right and left edgepoints of X are characterized as those points of X having an immediate predecessor and successor in X, respectively. Thus f maps such edgepoints onto the corresponding edgepoints for Y and can be extended linearly over the intervals of X'. It is routine to check that f, so extended, is an order isomorphism of $[0,1]$ onto itself.

Remark 3.4.7. With X, Y, and f, as above, construct μ and ν, measures on X and Y, respectively, as in Theorem 3.4.5. (We shall refer to μ and the ordered basis of Theorem 3.4.5 as "the canonical measure and ordered basis for X.") Denoting by f, again, the order-isomorphic extension of f constructed above, f is a Lebesgue order isomorphism on $[0,1]$ if and only if it is on X, if and only if f carries the null sets of μ onto those of ν. In fact, since f is linear on each of the countable number of connected components of X', it is a Lebesgue order isomorphism on each of these; so that f is a Lebesgue order isomorphism on $[0,1]$, if and only if it is on X. Now, the μ and ν null sets in X can be described as those subsets containing no right edgepoints and whose Lebesgue measure is zero, so that f is a Lebesgue order isomorphism on X if and only if f carries μ-null sets onto ν-null sets.

THEOREM 3.4.8. *The canonical ordered bases $\mathfrak{F}_1$ and $\mathfrak{F}_2$ for the sets X and Y, respectively, are unitarily equivalent if and only if X and Y are*

Lebesgue order isomorphic. The hull class of an ordered basis consists of the images of any one of its members under Lebesgue order isomorphisms of $[0,1]$ *onto itself.*

Proof. The hulls of $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are $\{E_\lambda\}$ and $\{F_\lambda\}$, respectively, where E_λ and F_λ are multiplication by the characteristic functions of $[0,\lambda]\cap X$ and $[0,\lambda]\cap Y$ with λ in X and λ in Y, respectively. Let f be an order isomorphism of X and Y, and define φ by: $\varphi(E_\lambda)=F_{f(\lambda)}$. The diagonals $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathfrak{T}_1$ and $\mathfrak{T}_2$, respectively, contain strongly dense C^*-subalgebras $\mathfrak{A}_1$ and $\mathfrak{A}_2$, respectively, consisting of multiplications by continuous functions on X and Y, respectively. Let T_g be the operator in $\mathfrak{A}_1$ corresponding to multiplication by g, and define $\varphi(T_g)$ to be $T_{g\circ f^{-1}}$, the operator in $\mathfrak{A}_2$ corresponding to multiplication by $g\circ f^{-1}$. If φ on $\{E_\lambda\}$ can be implemented by a unitary transformation, the unitary equivalence induced on $\mathcal{A}_1$, when restricted to $\mathfrak{A}_1$, is the mapping, φ, just defined; for T_g is approximable to within ϵ in bound by some finite linear combinations of the E_λ's, and the unitary equivalence transforms this linear combination (as φ) into one in the F_λ's approximating $T_{g\circ f^{-1}}$ to within ϵ in bound. On the other hand, a unitary transformation which implements φ on $\mathfrak{A}_1$ carries E_λ onto $F_{f(\lambda)}$; for E_λ is the greatest lower bound in $\mathcal{A}_1$ of the multiplications corresponding to continuous functions which are 1 on $[0,\lambda]\cap X$ and lie between 0 and 1 on $X-[0,\lambda]$, and φ on $\mathfrak{A}_1$ transforms this set onto the corresponding set for $F_{f(\lambda)}$. Thus, φ on $\{E_\lambda\}$ is implemented by a unitary transformation if and only if φ on $\mathfrak{A}_1$ is; and this last obtains, if and only if the two representations, ψ_1 and ψ_2, of $C(X)$ defined by: $\psi_1(g)=T_g$, $\psi_2(g)=T_{g\circ f^{-1}}$, are unitarily equivalent. Since the weak closures of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are maximal abelian, φ can be unitarily implemented if and only if φ has an isomorphic extension to these weak closures, and this occurs if and only if ψ_1 and ψ_2 have the same null sets (cf. [4, Corollary 2.3.1], for example). The constant function, 1, is a separating vector for $\mathcal{A}_1$ and $\mathcal{A}_2$, whence the null sets are easily computed as the μ-null sets in X for ψ_1 and f^{-1} of the ν-null sets in Y for ψ_2. By symmetry, we conclude that φ on $\{E_\lambda\}$ is implemented by a unitary transformation if and only if f carries the μ-null sets onto the ν-null sets.

Suppose now that $\mathfrak{T}$ is an ordered basis with hulls $\{E_\alpha\}$, diagonal $\mathcal{A}$, and x is a separating vector for $\mathcal{A}$. Let X be $\{(E_\alpha x,x)\}$, so that $X\in h(\mathfrak{T})$; and let $\mathfrak{T}_1$ be the canonical ordered basis for X. From Theorem 3.4.5, $X\in h(\mathfrak{T}_1)$; whence $\mathfrak{T}_1$ and $\mathfrak{T}$ are unitarily equivalent, by Theorem 3.4.4. If Y is a closed subset of $[0,1]$ containing 0 and 1, and $\mathfrak{T}_2$ is the canonical ordered basis for Y, then $\mathfrak{T}_2$ is unitarily equivalent to $\mathfrak{T}$ if and only if $\mathfrak{T}_2$

is unitarily equivalent to $\mathcal{J}_1$—which occurs if and only if there is a Lebesgue order isomorphism of X onto Y. On the other hand, $\mathcal{J}_2$ is unitarily equivalent to $\mathcal{J}$ if and only if Y is in the hull class of $\mathcal{J}$. Thus, $h(\mathcal{J})$ consists of the images of X under Lebesgue order isomorphisms of $[0,1]$ onto itself.

3.5. Special cases and examples. While the preceding section gives a fairly complete account from the general viewpoint, there is still much to be done regarding special cases. By virtue of the results of that section, the remaining problem can be recast as that of finding detailed information concerning special hull classes. We have not ruled out the possibility that "order isomorphism" will suffice in place of "Lebesgue order isomorphism" for the classification of hull classes. (We shall do so in this section.) The example of § 3.3 shows us that an order isomorphism need not be Lebesgue, but it is conceivable that its existence guarantees the existence of one which is Lebesgue. In fact, Theorem 3.2.1 shows that this is so in the totally-atomic case (in this instance, the order isomorphism itself is Lebesgue). In addition, the hull class of a non-atomic ordered basis consists of $[0,1]$, whence an order isomorphism guarantees the existence of one which is Lebesgue (e.g., the identity mapping), if one of the ordered bases is assumed non-atomic. Even after we have noted that this phenomenon does not hold in general, we may still inquire into the classification of hull classes by means of "order isomorphism" and special properties of sets—e.g. an order isomorphism class of measure 0 sets may form a total hull class (cf. Remark 3.5.9).

We begin with the description of a construction closely akin to the construction of the canonical measure for X.

LEMMA 3.5.1. *If X is a closed subset of $[0,1]$ containing 0 and 1, $I_1, I_2, \cdots$, the distinct connected components of $[0,1]-X$, r_k and l_k are the right and left endpoints of I_k, respectively, then the mapping θ of X into $[0,1]$ defined by:*

$$\theta(a)=a-\sum_{r_k\leqq a} m(I_k),$$

is continuous, order preserving, and identifies a and b if and only if $m([a,b]\cap X)=0$. The image, Z, of θ is a closed interval.

Proof. Clearly, $|\theta(a)-\theta(b)|\leqq|a-b|$, so that θ is continuous and Z is compact. If a and b are in X and $a<r_k\leqq b$, then $[a,b]$ contains I_k; so that

$$\theta(b)-\theta(a)=b-a-\sum_{a<r_k\leqq b} m(I_k)=m([a,b]\cap X).$$

Thus, $\theta(a)\leqq\theta(b)$, and $\theta(a)=\theta(b)$ if and only if $m([a,b]\cap X)=0$.

Let a' be $\sup\{a : a \in X \text{ and } \theta(a) \leqq c\}$, and b' be $\inf\{b : b \in X \text{ and } c \leqq \theta(b)\}$. Since X is closed, a' and b' lie in X; and since θ is continuous, $\theta(a') \leqq c \leqq \theta(b')$. However, $[a', b'] \cap X$ contains just a' and b' and, so, has measure 0. Thus $\theta(a') = \theta(b') = c$, and $c \in Z$. It follows that Z is a closed interval.

The set, Z, consists of 0 alone if and only if $\theta(1) - \theta(0) = 0$—which occurs if and only if $m([0,1] \cap X) = m(X) = 0$, i.e., $\sum m(I_k) = 1$. If $\mathcal{F}$ is an ordered basis with diagonal $\mathfrak{A}$, X in $h(\mathcal{F})$ corresponding to the separating vector, x, and E_k is the minimal projection in $\mathfrak{A}$ with $(h(E_k)x, x) = r_k$; then $l_k = ([h(E_k) - E_k]x, x)$, so that $m(I_k) = (E_k x, x)$, and $((\sum E_k)x, x) = \sum m(I_k) = 1$. This is equivalent to $\mathfrak{A}$ being totally atomic.

LEMMA 3.5.2. *If X and Y are closed, measure-zero subsets of $[0,1]$ with infima a and c and suprema b and d, respectively, and f is an order isomorphism of the right edgepoints of X in $[a,b]$ onto those of Y in $[c,d]$, there is a Lebesgue order-isomorphic extension of f mapping $[a,b]$ onto $[c,d]$ and carrying X onto Y.*

Proof. Clearly, it suffices to consider the case where $a = c = 0$ and $b = d = 1$. Assuming this, let μ and ν be the canonical measures and $\mathcal{F}$ and $\mathcal{S}$ the canonical ordered bases for X and Y, respectively. From the comment preceding this lemma, $\mathcal{F}$ and $\mathcal{S}$ have totally-atomic diagonals. From Remark 3.4.3, the hulls of the minimal projections in $\mathcal{F}$ and $\mathcal{S}$ are order isomorphic with the right edgepoints of X and Y, respectively, and, so, to each other. According to Theorem 3.2.1, this isomorphism can be implemented by a unitary transformation carrying $\mathcal{F}$ onto $\mathcal{S}$. This unitary transformation carries the hulls of $\mathcal{F}$ onto those of $\mathcal{S}$ order isomorphically. By means of the order isomorphisms of X with the hulls of $\mathcal{F}$ and Y with those of $\mathcal{S}$, we arrive at a Lebesgue order isomorphism (recall that $0 = m(X) = m(Y)$) of X onto Y extending f. Remarks 3.4.6, 3.4.7 imply the existence of the desired extension of f to $[0,1]$.

If $\mathfrak{A}$ is not totally atomic, we may normalize θ by composing it with multiplication by $1/m(Z)\,(=1/m(X))$ to get a mapping of X onto $[0,1]$ with the properties of θ noted in Lemma 3.5.1. We denote this new mapping by θ_X (in the case where $\mathfrak{A}$ is not totally atomic). Writing p_k for $\theta_X(r_k)$ and $o(p_k)$ for the order type of the set of right edgepoints of X in $\theta_X^{-1}(p_k)$ (so that $o(p_k)$ is some denumerable total-order type), we shall say that X has a uniform edge of type τ if each $o(p_k) = \tau$. The construct consisting of $[0,1]$, the points p_k, and their associated order types, will be called "a point ordered interval" and the set of points $\{p_k\}$ its "atoms."

LEMMA 3.5.3. *If X and Y are closed subsets of* $[0,1]$ *containing* 0 *and* 1, $m(X) \neq 0 \neq m(Y)$, *f is a Lebesgue order isomorphism of X onto Y, and g is defined by* $g(p) = \theta_Y(f(\theta_X^{-1}(p)))$, *for each p in* $[0,1]$, *then g is a Lebesgue order isomorphism of* $[0,1]$ *onto itself,* $\{g(p_k)\}$ *are the atoms of the point ordered interval for Y, where* $\{p_k\}$ *are those of the interval for X, and* $o(p_k) = o(g(p_k))$. *If a mapping such as g is given, then there is a mapping, f, such that* $g(p) = \theta_Y(f(\theta_X^{-1}(p)))$, *for each p in* $[0,1]$, *which is a Lebesgue order isomorphism of X onto Y.*

Proof. As defined, g is not obviously single-valued, since $\theta_X^{-1}(p)$ may contain more than one point. However, if a and b are in $\theta_X^{-1}(p)$, then $m([a,b] \cap X) = 0$, from Lemma 3.5.1; and

$$0 = m(f([a,b] \cap X)) = m([f(a),f(b)] \cap Y),$$

since f is a Lebesgue order isomorphism, whence $\theta_Y(f(a)) = \theta_Y(f(b))$, again from Lemma 3.5.1. Thus g is single-valued. Since θ_Y, θ_X, and f are order preserving, g is. With p_k equal to $\theta_X(r_k)$ and $f(r_k)$ a right edgepoint of Y (since f is an order isomorphism), we have that $g(p_k)$ is $\theta_Y(f(r_k))$, an atom of the point ordered interval for Y.

In the proof of Lemma 3.5.1, we noted that $\theta(b) - \theta(a) = m([a,b] \cap X)$, so that $\theta_X(b) - \theta_X(a) = m([a,b] \cap X)/m(X)$. Thus, $m(\theta_X^{-1}([p,q])) = (q-p)m(X)$, and $m(\theta_X^{-1}(S)) = m(S)m(X)$, for each open subset, S, of $[0,1]$. By regularity of m, if $m(S') = 0$, then $m(\theta_X^{-1}(S')) = 0$. On the other hand, from the fact that $|\theta_Y(b) - \theta_Y(a)| \leqq |b-a|/m(Y)$, we conclude that if $m(Y_0) = 0$, for a subset, Y_0, of Y, then $m(\theta_Y(Y_0)) = 0$. Thus, if $m(S) = 0$, for S a subset of $[0,1]$, $m(\theta_X^{-1}(S)) = 0$, $m(f(\theta_X^{-1}(S))) = 0$, and $0 = m(\theta_Y(f(\theta_X^{-1}(S)))) = m(g(S))$.

What we have proved for g holds as well for the mapping $\theta_X f^{-1} \theta_Y^{-1}$, which is clearly g^{-1}. It follows that g is a Lebesgue order isomorphism of $[0,1]$ onto itself mapping atoms (for X) onto atoms (for Y). When we note that $\theta_Y^{-1}(g(p_k)) = f(\theta_X^{-1}(p_k))$ and f maps the right edgepoints of X onto those of Y (order isomorphically), we see that the right edgepoints of $\theta_Y^{-1}(g(p_k))$ are order isomorphic to those of $\theta_X^{-1}(p_k)$, whence $o(p_k) = o(g(p_k))$.

Suppose, now, that g is given with the properties described above. We show that there is a Lebesgue order isomorphism, f, of X onto Y such that $g = \theta_Y f \theta_X^{-1}$. We note first that if $\theta_X^{-1}(p)$ contains more than one point then p is an atom for X (i.e. p is some p_k). In fact, $\theta_X^{-1}(p) = [a,b] \cap X$, where $a = \inf\{a' : a' \in \theta_X^{-1}(p)\}$, $b = \sup\{b' : b' \in \theta_X^{-1}(p)\}$, a and b are in $\theta_X^{-1}(p)$, and $m([a,b] \cap X) = 0$ (cf. Lemma 3.5.1). Thus,

$$\sum_{I_k \subsetneq [a,b]} m(I_k) = b - a \neq 0;$$

and in particular, there is some r_k in $[a,b]$. Of course, $\theta_X(r_k)=p=p_k$. If p is not an atom for X, then by assumption, $g(p)$ is not an atom for Y, whence $\theta_X^{-1}(p)$ and $\theta_Y^{-1}(g(p))$ each consist of a single point. Define $f(\theta_X^{-1}(p))$ to be $\theta_Y^{-1}(g(p))$, for such points, p. By assumption on g, $o(p_k)=o(g(p_k))$, which guarantees an order isomorphism of the right edgepoints of X in $\theta_X^{-1}(p_k)$ onto those of Y in $\theta_Y^{-1}(g(p_k))$. If $\theta_X^{-1}(p_k)=[a,b]\cap X$ and $\theta_Y^{-1}(g(p_k))=[c,d]\cap Y$, the right edgepoints of $\theta_X^{-1}(p_k)$ in $[a,b]$ and $\theta_Y^{-1}(g(p_k))$ in $[c,d]$ are those of X which lie in $[a,b]$ and those of Y which lie in $[c,d]$, respectively. This is clear with the possible exception of the points a and c. However, a cannot be a right edgepoint of X for then the corresponding left edgepoint would lie in $\theta_X^{-1}(p_k)$ but not in $[a,b]$. Similarly, c is not a right edgepoint of Y. It follows from Lemma 3.5.2 that there is a Lebesgue order isomorphism of $[a,b]\cap X$ onto $[c,d]\cap Y$. We define f on $[a,b]\cap X$ to be this isomorphism. Clearly, $g=\theta_Y f\theta_X^{-1}$; and f is an order isomorphism. If $m(X_0)=0$ for a subset, X_0, of X, then $m(\theta_X(X_0))=0$ and $m(\theta_Y^{-1}(g(\theta_X(X_0))))=0$, from the first part of this proof, since $m(g(\theta_X(X_0)))=0$. Now $f(X_0)$ is contained in $\theta_Y^{-1}(g(\theta_X(X_0)))$; whence $m(f(X_0))=0$, and f is a Lebesgue order isomorphism.

By virtue of the preceding lemma, the study of hull classes is equivalent to the study of denumerable subsets of $[0,1]$ each point of which has a denumerable total-order type associated with it, under Lebesgue order isomorphisms of $[0,1]$ onto itself. The next lemma shows that all possibilities occur as point ordered intervals.

LEMMA 3.5.4. *If $\{p_k\}$ is a denumerable subset of $[0,1]$ and $\{\tau_k\}$ is a set of denumerable total-order types, then there is a closed subset of $[0,1]$ containing 0 and 1 whose point ordered interval has $\{p_k\}$ as its atoms and τ_k as the order type associated with p_k.*

Proof. We begin by showing that each non-zero interval contains a closed subset, with the interval endpoints as members, having measure 0, and whose right edgepoints have some preassigned denumerable order type, τ. If τ is the order type of a non-null finite set, this result is clear. We assume that τ is the order type of some infinite (denumerable) set, $\{a_j\}_{j=1,2,\cdots}$. We may assume, in addition, that the interval in question is $[0,1]$ (translating and multiplying by a suitable scalar). For each positive integer, k, let b_k be $\sum_n 2^{-k_n}$, where $\{a_{k_n}\}$ is the subset of $\{a_j\}$ consisting of points not exceeding a_k; and let I_k be the open interval of length 2^{-k} with b_k as right endpoint. If $a_j<a_k$ then $b_k-b_j\geqq 2^{-k}$, by construction; whence I_k and I_j are disjoint.

Let X be the complement in $[0,1]$ of the union of the intervals, I_k. Clearly, X is closed, contains 0 and 1, has $\{b_k\}$ as its set of right edgepoints, and has measure 0 (since $\sum_{k=1}^{\infty} m(I_k) = \sum_{k=1}^{\infty} 2^{-k} = 1$). The comment establishing the disjointness of I_j and I_k, when $j \neq k$, also establishes the fact that the correspondence $a_j \to b_j$ is an order isomorphism of $\{a_j\}$ with $\{b_j\}$. Thus, the right edgepoints of X have order type τ.

If $\{p_k\}$ is a finite subset of $[0,1]$, there is no difficulty in establishing the conclusion of this lemma (when the result just proved is employed). We assume that $\{p_k\}$ is an infinite set. Let c_k be $(\sum_n 2^{-(k_n+1)}) + p_k/2$, where $\{p_{k_n}\}$ is the subset of $\{p_j\}$ consisting of those numbers which do not exceed p_k; and let J_k be the closed interval of length $2^{-(k+1)}$ with c_k as right endpoint (and d_k as left endpoint). Let X_k be a closed, measure 0 subset of J_k containing c_k and d_k and having right edgepoints in J_k with τ_k as order type. The complement, X, of $\bigcup_k (J_k - X_k)$ in $[0,1]$ is a closed subset of $[0,1]$ containing 0 and 1 (since neither of these points lies in $J_k - X_k$, $k = 1, \cdots$). If $p_j < p_k$, then $c_k - c_j \geqq 2^{-(k+1)} + (p_k - p_j)/2$, so that J_k and J_j are disjoint. The components of the complement of X in $[0,1]$ are, therefore, the aggregate of the components of the complement of each X_k in J_k, so that the right edgepoints of X in $[0,1]$ are those of each X_k in J_k. (Note that d_k is not a right edgepoint of X, by disjointness of J_k and J_j, with $j \neq k$.) Now, $\theta(1) = 1 - \sum_{j=1}^{\infty} 2^{-(j+1)} = 1/2$, so that $\theta_X = 2\theta$; and $\theta_X(c_k) = 2(p_k/2) = p_k$. Thus, each p_k is an atom for X; and since $m(X_k) = 0$, $\theta_X(X_k) = p_k$, so that $\{p_k\}$ is the set of atoms for X. The right edgepoints of X which θ_X maps onto p_k are those of X_k in J_k, and therefore has τ_k as order type; for any other right edgepoint of X lies in some X_j, with $j \neq k$, and $\theta_X(X_j) = p_j \neq p_k$. Thus X has a point ordered interval with $\{p_k\}$ as its set of atoms, each p_k associated with the denumerable, total-order type, τ_k.

The lemma which follows provides the key to the description of a special family of hull classes.

LEMMA 3.5.5. *If R and S are dense denumerable subsets of $[0,1]$ both containing 0 and 1, and m, M are numbers such that $0 < m < 1 < M$, then there exists a homeomorphism, f, of $[0,1]$ onto itself such that:*

(i) $f(0) = 0$, $f(1) = 1$;

(ii) *f maps R onto S;*

(iii) $m(x-y) \leqq f(x) - f(y) \leqq M(x-y)$, *for each* $x \geqq y$ *in* $[0,1]$.

Proof. If we construct a mapping, g, of R onto S satisfying (i) and (iii) for points x, y in R, then g is uniformly continuous and so has a unique continuous extension, f, to $[0,1]$. By density of R and continuity of f, (iii) holds; whence f is a homeomorphism of $[0,1]$ onto itself satisfying (i), (ii), (iii).

To construct such a g, we begin by enumerating the sets R and S as $r_1 = (0)$, $r_2(=1), r_3, \cdots$ and $s_1(=0), s_2(=1), s_3, \cdots$, respectively. When we refer to "the points of $\{a_1, \cdots, a_n\}$ adjacent to a", we shall mean those points a_j, a_k such that $a_j < a < a_k$ and $a_j < a_h < a_k$, for $1 \leqq h \leqq n$, implies $a_h = a$. We write (r, s, r', s') in place of the inequalities, $m < \dfrac{s' - s}{r' - r} < M$.

Define "an s link of length h," for s in S, to be a set of ordered pairs $\{(r_1, s_{n_1}), (r_2, s_{n_2}), \cdots, (r_h, s_{n_h})\}$ such that $(r_j, s_{n_j}, r_k, s_{n_k})$, for $j \neq k$; j, $k = 1, \cdots, h$; and $s_{n_h} = s$.

If $L = \{(r_1, s_{n_1}), \cdots, (r_k, s_{n_k})\}$ is an s_{n_h} link, $h > k$, and s_{n_h} is such that $(r_{n_j}, s_{n_j}, r_h, s_{n_h})$, for $j = 1, \cdots, k$, then there is an s_{n_h} link of length h of which L is a subset. In fact, with $k \leqq a < b < h$, suppose that we have found s_{n_a} such that $(r_j, s_{n_j}, r_a, s_{n_a})$, for $j = h, 1, 2, \cdots, a-1$. Let $r_{j'}$ and $r_{j''}$ be the points of $\{r_h, r_1, r_2, \cdots, r_{b-1}\}$ adjacent to r_b. The set of points, x, for which $(r_b, x, r_{j'}, s_{n_{j'}})$ and $(r_b, x, r_{j''}, s_{n_{j''}})$, is a non-null open set, since $(r_{j'}, s_{n_{j'}}, r_{j''}, s_{n_{j''}})$; so that it contains elements of the dense set, S. Let s_{n_b} be that element of lowest index. Since $r_{j'}$, $r_{j''}$ are adjacent to r_b and $(r_{j'}, s_{n_{j'}}, r_b, s_{n_b})$, $(r_{j''}, s_{n_{j''}}, r_b, s_{n_b})$, we have $(r_j, s_{n_j}, r_b, s_{n_b})$, for each $j = h, 1, \cdots, b-1$. In this way, we construct an s_{n_h} link containing L.

Suppose, now, that we have defined $g(r_1), \cdots, g(r_{n-1})$ so that $\{(r_1, g(r_1)), \cdots, (r_{n-1}, g(r_{n-1}))\}$ $(= L)$ is a link, and $g(r_1) = 0$, $g(r_2) = 1$. Let s be the element of $S - \{g(r_1), \cdots, g(r_{n-1})\}$ with least index; and let S_n be the set of elements of S paired with r_n in those s links containing L which have minimal length. We shall define $g(r_n)$ to be the element of S_n with least index, so that $\{(r_1, g(r_1)), \cdots, (r_n, g(r_n))\}$ is a link; but first, we must show that S_n is not empty. For this, it will suffice to prove that some s link containing L exists. Let $g(r_{j'})$ and $g(r_{j''})$ be the elements of $\{g(r_1), \cdots, g(r_{n-1})\}$ adjacent to s. As in the preceding paragraph (by density of R), there is an element r_h, of R such that $(r_{j'}, g(r_{j'}), r_h, s)$ and $(r_{j''}, g(r_{j''}), r_h, s)$; whence $(r_j, g(r_j), r_h, s)$, for all $j = 1, \cdots, n-1$. From our preceding comments, we conclude that there is an s link containing L.

If g does not map R onto S, let s be the element of S not in the range of g with least index; and suppose that each element of S with index less

than that of s is contained among $g(r_1), g(r_2), \cdots, g(r_k)$. From the above, there is some s link, of length h, let us say, containing

$$\{(r_1, g(r_1)), \cdots, (r_k, g(r_k))\}.$$

Let $\{(r_1, g(r_1)), \cdots, (r_j, g(r_j)), (r_{j+1}, s_{n_{j+1}}), \cdots, (r_{h'}, s)\}$ be an s link of length not exceeding h for which j is maximal (since s is not in the range of g, there is some first $s_{n_{j+1}}$ which is not $g(r_{j+1})$). Certainly then $j \geqq k$, so that s is the element of $S - \{g(r_1), \cdots, g(r_j)\}$ with least index. By definition, $g(r_{j+1})$ occurs with r_{j+1} in an s link of length not exceeding h' and containing $\{(r_1, g(r_1)), \cdots, (r_j, g(r_j))\}$. This contradicts the maximal property of j; whence g maps R onto S, and the proof is complete.

Remark 3.5.6. The result of the preceding lemma is valid in the case where R and S both contain or both do not contain 0 or 1 (as can be seen by adjoining 0 or 1 to the sets—whichever is appropriate).

If X is a closed subset of $[0,1]$ containing 0 and 1, we shall say that two points, a and b, of X are "equivalent in X" when $m([a,b] \cap X) = 0$ (equivalently, when $\theta(a) = \theta(b)$—cf. Lemma 3.5.1).

THEOREM 3.5.7. *If τ is a denumerable order type, the family, $\mathfrak{F}$, of closed subsets, X, of $[0,1]$ containing 0 and 1, which are nowhere dense, have uniform edge of type τ, have non-zero Lebesgue measure, and for which 0 and 1 are equivalent in X to right edgepoints of X is the hull class of some ordered basis. The family of subsets having the same properties and for which 0 or 1 are not equivalent to a right edgepoint is also a hull class.*

Proof. If $X \in \mathfrak{F}$ and f is a Lebesgue order isomorphism of $[0,1]$ onto itself, then clearly, $f(X) \in \mathfrak{F}$. Thus, $\mathfrak{F}$ contains the hull class determined by X.

With $0 \leqq p < q \leqq 1$, $\theta^{-1}([p,q]) = [a,b] \cap X$, since θ is order preserving. Now X is nowhere dense, so that $[a,b] - X$ is non-null. Hence, a right edgepoint of X lies in $[a,b]$ and some atom for X lies in $[p,q]$. The set of atoms for each set of $\mathfrak{F}$ is everywhere dense in $[0,1]$ and contains 0 and 1 (in the case of the first statement of this theorem). According to Lemma 3.5.5, there is a Lebesgue order isomorphism of $[0,1]$ onto itself carrying the atoms for X onto those for Y, where X and Y are sets in $\mathfrak{F}$, and necessarily preserving the order types associated with these atoms, since by hypothesis, τ is associated with all the atoms for X and Y. Lemma 3.5.3 now applies, and we conclude that there is a Lebesgue order isomorphism of X and Y. Thus $\mathfrak{F}$ constitutes a complete hull family.

Remark 3.5.8. We note in the foregoing theorem that X being nowhere dense implies that its atoms are everywhere dense. Suppose that we are given that the atoms for X are everywhere dense in $[0,1]$. By definition of θ_X, an open interval in X maps in a one-one manner onto an interval in $[0,1]$ which contains no atoms for X. Thus X contains no open intervals, and being closed, X is nowhere dense.

Remark 3.5.9. The density of the atoms for X in $[0,1]$ is equivalent to their closure having measure 1. The unitary equivalence class of ordered bases corresponding to the hull class of Theorem 3.5.7 can also be described as the set of ordered bases whose point ordered intervals have a set of atoms whose closure has measure 1, contains 0 and 1 (similarly for the other three cases), and each atom is associated with a fixed order type, τ. In this framework, we can state a similar result for the case where the closure of the set of atoms has measure 0. In fact, the family of all denumerable subsets of $[0,1]$ equivalent to a given one under order isomorphisms of $[0,1]$ onto itself, all of whose closures have measure 0, and associated with each point of which is a fixed denumerable total-order type, τ, constitutes the family of point ordered intervals corresponding to a unitary equivalence class of ordered bases. Clearly, Lebesgue order isomorphisms of $[0,1]$ onto itself leave this family invariant; while the order isomorphism between the closures of two sets of the family is Lebesgue in these sets (since these closures have measure 0) and can, therefore, be extended to a Lebesgue order isomorphism of $[0,1]$ onto itself (cf. Remark 3.4.7). The order isomorphism of the sets of atoms, themselves, would not be sufficient; for unlike the situation of Remark 3.4.6, an order isomorphism between subsets of $[0,1]$ which are not closed need not be extendable to an order isomorphism of $[0,1]$ onto itself (and so, from Remark 3.4.6, not extendable to their closures). Indeed, the sets, $A=\{1/2-1/n, 1/2+1/n\}_{n=2,3,\cdots}$ and $B=\{1/4-1/n, 3/4+1/n\}_{n=4,5,\cdots}$ are order isomorphic but their closures in $[0,1]$ are not (the closure of A has a point, $1/2$, without immediate predecessor or successor, while the closure of B has no such point).

The examples which follow indicate some of the limitations to the possibility of simple characterization of hull classes. Making use of a non-uniform edge, the next example gives us our first instance of order isomorphic sets which do not belong to the same hull class.

Example 3.5.10. Let g be a homeomorphism of $[0,1]$ onto itself which carries some set of measure 0 onto a set of measure different from 0 (such as f, described in § 3.3). Let $r_1, r_2, \cdots$ be an enumeration of the rationals

in $[0,1]$, and let τ_k be the total-order type corresponding to the totally-ordered sets with k elements (a finite set). If X and Y are closed subsets of $[0,1]$ which have $\{r_k, \tau_k\}$ and $\{g(r_k), \tau_k\}$ as point ordered intervals (cf. Lemma 3.5.4) then there is an order isomorphism, f, of X onto Y such that $g = \theta_Y f \theta_X^{-1}$, from Lemma 3.5.3. (Note that the second part of the proof of Lemma 3.5.3 applies to order isomorphisms, g, which are not Lebesgue to give order isomorphisms, f, which are not Lebesgue.) If there is a Lebesgue order isomorphism, f_0, of X onto Y, then there is a Lebesgue order isomorphism, g_0, of $[0,1]$ onto itself carrying r_k onto $g(r_{j_k})$, $k = 1, 2, \cdots$, such that $\tau_k = \tau_{j_k}$ and $g_0 = \theta_Y f_0 \theta_X^{-1}$, from Lemma 3.5.3. Since τ_k and τ_{j_k} are the order types of finite sets with k and j_k elements, respectively, $k = j_k$; whence $g_0(r_k) = g(r_{j_k}) = g(r_k)$. Both g and g_0 are continuous and $\{r_k\}$ is dense in $[0,1]$, so that $g = g_0$. But g_0 is Lebesgue and g is not. Thus X and Y, though order isomorphic, do not lie in the same hull class.

In the foregoing example, special use was made of the fact that the edge of the closed sets involved was not uniform. The next example describes a case in which two order isomorphic sets with uniform edge (of any type we wish) do not belong to the same hull class and whose sets of atoms have closures with the same measure (not 0 or 1, of course, in view of Theorem 3.5.7, and Remark 3.5.9).

Example 3.5.11. Let C_1 and C_2 be dense denumerable subsets of Cantor sets of measures 0 and 1/4, respectively, in $[0, 1/2]$, both C_1 and C_2 containing 1/2; and let g' be a homeomorphism of $[0, 1/2]$ onto itself carrying C_1 onto C_2 (multiply by 1/2 in the preceding example). Let D_1 be a dense denumerable subset of $[5/8, 1]$ containing 5/8; and let D_2 be its image under the linear order preserving homeomorphism, g'', of $[5/8, 1]$ onto $[7/8, 1]$. The mapping, g, defined as g' on $[0, 1/2]$, g'' on $[5/8, 1]$, and linear from $[1/2, 5/8]$ to $[1/2, 7/8]$, is an order isomorphism of $[0,1]$ onto itself carrying $S_1 (= C_1 \cup D_1)$ onto $S_2 (= C_2 \cup D_2)$. Note that the closures of S_1 and S_2 have measure 3/8. If h is an order isomorphism of S_1 onto S_2, then $h(1/2)$ is a point of S_2 with an immediate successor and, hence, a point of C_2. If $h(1/2) \neq 1/2$, then some point of D_1 other than 5/8 maps onto 7/8. However, no point of D_1 other than 5/8 has an immediate predecessor, while 7/8 does. Thus $h(1/2) = 1/2$, $h(5/8) = 7/8$, $h(C_1) = C_2$, and $h(D_1) = D_2$. Since the closure of C_1 has zero measure and that of C_2 does not, no Lebesgue order isomorphism of $[0,1]$ onto itself carries S_1 onto S_2. From Lemma 3.5.3, X and Y, closed subsets of $[0,1]$ whose point ordered intervals have S_1 and S_2 for their sets of atoms, respectively, are not in the same hull class. However,

there is an order isomorphism, viz. g, of $[0,1]$ onto itself carrying S_1 onto S_2, and again, from Lemma 3.5.3, X and Y are order isomorphic.

Added February 3, 1960: Question 2.4.1 has a negative answer as can be seen with the aid of a result of W. F. Donaghue, "The lattice of invariant subspaces of a completely continuous quasinilpotent transformation," *Pacific Journal of Mathematics*, vol. 7 (1957), pp. 1031-1035.

COLUMBIA UNIVERSITY
AND
MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien*, Paris, 1957.

[2] B. Fuglede, "A commutativity theorem for normal operators," *Proceedings of the National Academy of Sciences (USA)*, vol. 36 (1950), pp. 35-40.

[3] P. Halmos, *Measure theory*, New York, 1950.

[4] R. Kadison, "Unitary invariants for representations of operator algebras," *Annals of Mathematics*, vol. 66 (1957), pp. 304-379.

[5] ——— and I. Singer, "Extensions of pure states," *American Journal of Mathematics*, vol. 81 (1959), pp. 383-400.

[6] F. Murray and J. von Neumann, "Rings of operators," *Annals of Mathematics*, vol. 37 (1936), pp. 116-229.

FINITE DIMENSIONAL PERTURBATIONS IN BANACH SPACES.* [1]

By S. R. Foguel.

0. Introduction. It was observed in [2] that two differential operators on a finite interval, which arise from a given differential equation by two sets of boundary conditions, have resolvent operators which differ by a finite dimensional operator. In this paper we study the following situation: Let X be a Banach space, T a bounded operator and S a bounded operator of the form $Sx = x_0^*(x)x_0$, where $x_0 \in X$, $x_0^* \in X^*$. We shall be concerned with the one dimensional perturbation of T: $T + S$.

Thus in the case of differential operators, one set of boundary conditions will be arbitrary, while the second set will be at an end point of the interval. The second operator will, then, be a quasi nilpotent and the first will be a finite dimensional perturbation of a quasi nilpotent.

1. The resolvent of $T + S$. Let $\sigma = \sigma(T)$ and $R_\mu = R_\mu(T)$ (the spectrum and resolvent of T). If $\mu \notin \sigma$ is a complex number such that $\| R_\mu S \| < 1$, then $(\mu - T - S)^{-1}x = (I - R_\mu S)^{-1}R_\mu x = \sum_{n=0}^{\infty} (R_\mu S)^n R_\mu x$ is

$$R_\mu x + (x_0^* R_\mu x) R_\mu x_0 + \cdots + (x_0^* R_\mu x)(x_0^* R_\mu x_0)^{n-1} R_\mu x_0 + \cdots$$
$$= R_\mu x + (1 - x_0^* R_\mu x_0)^{-1}(x_0^* R_\mu x) R_\mu x_0.$$

Notation: Let the functions ϕ and ψ be defined by $\phi(\mu) = 1 - x_0^* R_\mu x_0$, $\psi(\mu; x) = (\phi(\mu))^{-1}(x_0^* R_\mu x)\ R_\mu x_0$.

Theorem 1.1. *If $\mu \notin \sigma$ and $x_0^* R_\mu x_0 \neq 1$, then $\mu \notin \sigma(T + S)$ and*

$$(\mu - T - S)^{-1}x = R_\mu x + \psi(\mu; x). \tag{1.1}$$

If $\mu \notin \sigma$ and $\phi(\mu) = 0$, then μ is an eigenvalue of $T + S$ with a single eigenvector $R_\mu x_0$.

Proof. Whenever the right side of equation (1.1) is defined, then

[1] This work was partially supported by the National Science Foundation.

$$
\begin{aligned}
(\mu-T-S)&(R_\mu x+\psi(\mu;x)) \\
&= x - SR_\mu x + \phi(\mu)^{-1}x_0^*R_\mu x(x_0 - SR_\mu x_0) \\
&= x - (x_0^*R_\mu x)x_0 + \phi(\mu)^{-1}x_0^*R_\mu x(x_0 - (x_0^*R_\mu x_0))x_0 = x.
\end{aligned}
$$

Also

$$
\begin{aligned}
R_\mu&(\mu-T-S)x + \phi(\mu)^{-1}(x_0^*R_\mu(\mu-T-S)x)R_\mu x_0 \\
&= x - (x_0^*x)R_\mu x_0 + \phi(\mu)^{-1}[(x_0^*x) - (x_0^*x)(x_0^*R_\mu x_0)]R_\mu x_0 = x.
\end{aligned}
$$

If $\phi(\mu)=0$, then

$$(\mu-T-S)R_\mu x_0 = x_0 - SR_\mu x_0 = x_0 - (x_0^*R_\mu x_0)x_0 = 0.$$

Conversely, if $(\mu-T-S)x=0$, then $x=R_\mu Sx=(x_0^*x)R_\mu x_0$.

Remark. The above theorem characterizes $\sigma(T+S)-\sigma(T)$. The function $x_0^*R_\mu x_0$ is analytic outside of $\sigma(T)$ and vanishes at infinity. We shall assume that $\sigma(T)$ is connected; then $\sigma(T+S)-\sigma(T)$ consists of a sequence of points which has limit points, if any, in $\sigma(T)$.

In [2] a result equivalent to the second part of Theorem 1.1 is proved for self adjoint operators.

2. Computations of $f(S+T)$. Let f be a function analytic for $|\mu)|<|\sigma(T+S)|+2\epsilon$, $\epsilon>0$. Thus

$$f(\mu)=\sum_{n=0}^{\infty}a_n\mu^n \text{ in this domain.}$$

Then

$$f(T+S)x=\frac{1}{2\pi i}\int_c f(\mu)(\mu-T-S)^{-1}x\,d\mu,$$

where c: $|\mu|=|\sigma(T+S)|+\epsilon$ (See [1], Definition 2.6), or

$$f(T+S)x=f(T)x+\frac{1}{2\pi i}\int_c f(\mu)\psi(\mu;x)\,d\mu.$$

Note that $f(T+S)-f(T)$ is a compact operator. Now

$$(1-x_0^*R_\mu x_0)^{-1}x_0^*R_\mu x=\sum_{n=0}^{\infty}b_n/\mu^{n+1}, \quad |\mu|>|\sigma(T+S)|,$$

$$R_\mu x_0=\sum_{n=0}^{\infty}T^n x_0/\mu^{n+1}, \qquad |\mu|>|\sigma(T);$$

then $\frac{1}{2\pi i}\int_c f(\mu)\psi(\mu;x)\,d\mu$ is the coefficient of μ^{-1} in the function $f(\mu)\psi(\mu;x)$. But

$$\psi(\mu;x)=\mu^{-2}\sum_{n=0}^{\infty}(b_0T^nx_0+b_1T^{n-1}x_0+\cdots+b_nx_0)\mu^{-n}$$

and the coefficient of μ^{-1} is

$$\sum_{n=0}^{\infty}a_{n+1}(b_0T^nx_0+b_1T^{n-1}x_0+\cdots+b_nx_0)=\sum_{k=0}^{\infty}(\sum_{n=0}^{\infty}a_{n+k+1}b_n)T^kx_0.$$

3. Spectral behavior of $T+S$ at $\sigma(T+S)-\sigma(T)$. Let $\mu_0\notin\sigma$ and $\phi(\mu_0)=0$. Thus μ_0 is an isolated point of $\sigma(T+S)$. Let c be a circle around μ_0 which does not contain any other point of $\sigma(T+S)$. Then if

$$E=\frac{1}{2\pi i}\int_c(\mu-T-S)^{-1}\,d\mu$$

E is a projection and

$$T+S=(T+S)E+(T+S)(I-E)=(\mu_0-N)E+(K)(I-E),$$

where: The operator N on EX is a quasi nilpotent. The spectrum of the operator $T+S$ on $(I-E)X$ does not contain μ_0. (See [1], p. 196). Also $N=EN=NE$, $K=K(I-E)=(I-E)K$.

Now, the function R is analytical near μ_0; hence

$$Ex=\frac{1}{2\pi i}\int_c\psi(\mu;x)\,d\mu,$$

or Ex is the residue of the analytic function $\psi(\mu;x)$.

Let us divide the discussion into two parts.

Case 1. Let us assume here that

$$(d/d\mu)x_0{}^*R_\mu x_0\,|_{\mu=\mu_0}=-x_0{}^*R^2{}_{\mu_0}x_0\neq0.$$

In this case there is a simple pole to $\psi(\mu;x)$ and $N=0$ and

$$Ex=\lim_{\mu\to\mu_0}(\mu-\mu_0)\psi(\mu;x)=(x_0{}^*R_{\mu_0}x)R_{\mu_0}x_0/x_0{}^*R^2{}_{\mu_0}x_0.$$

This is a one dimensional projection whose norm is

$$\|E\|=(\|R_{\mu_0}x_0\|/|x_0{}^*R^2{}_{\mu_0}x_0|)\sup_{\|x\|=1}|x_0{}^*R_{\mu_0}x|$$
$$=\|R_{\mu_0}x_0\|\,\|R^*{}_{\mu_0}x_0{}^*\|/|x_0{}^*R^2{}_{\mu_0}x_0|.$$

Case 2. Let $\phi(\mu_0) = 0$ and

$$x_0{}^*R^2{}_{\mu_0}x_0 = \cdots = x_0{}^*R^n{}_{\mu_0}x_0 = 0 \neq x_0{}^*R^{n+1}{}_{\mu_0}x_0.$$

The operator $\psi(\mu)$ is one dimensional and thus compact. Therefore E is a compact operator too; the space EX is finite dimensional. Now N is defined on EX and therefore is a nilpotent. Let the dimension of EX be k. The operator N can be represented by the Jordan canonical form as

$$N = \begin{bmatrix} N_1 & 0 & 0 \\ 0 & N_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix}, \text{ where } N_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

It was shown in Theorem 1.1 that there is a unique solution to the equation $(\mu_0 - T - S)x = Nx = 0$. Hence $N = N_1$.

But then the pole of $T + S$ on EX at μ_0 is of order k: $k = n$, for $\phi(\mu)$ has a zero of order n while the numerator of $\psi(\mu)$ is different from zero at μ_0. Now let $y_m = R^m{}_{\mu_0}x_0$, $1 \leqq m \leqq n$; then

$$(3.1) \quad (\mu_0 - T - S)y_m = y_{m-1}, \quad n \geqq m \geqq 1, \quad (\mu_0 - T - S)y_1 = 0.$$

Therefore $y_m \in EX$ and $Ny_m = y_{m-1}$, where $y_0 = 0$. Now the vectors y_i are independent by equation (3.1) and hence generate EX. With respect to the basis y_i the operator N has the Jordan canonical form. Also there exist n functionals $y_1{}^*, \cdots, y_n{}^*$ such that

$$Ex = \sum_{i=1}^{n} y_i{}^*(x)R^i{}_{\mu_0}x_0 = \sum_{i=1}^{n} y_i{}^*(x)y_i.$$

In the rest of this section we compute the functionals $y_i{}^*$. The operator E is a projection, hence

$$E^2x = \sum_{i,j=1}^{n} y_i{}^*(x)y_j{}^*(y_i)y_j = Ex = \sum_{i=1}^{n} y_i{}^*(x)y_i$$

But the vectors y_i are independent, so

$$y_k{}^*x = y_k{}^*(\sum_{i=1}^{n} y_i{}^*(x)y_i) = y_k{}^*Ex,$$

or

$$(3.2) \qquad y_k{}^*(x - Ex) = 0.$$

Also

$$y_j = Ey_j = \sum_{i=1}^{n} y_i{}^*(y_j)y_i.$$

Thus

$$y_i^* y_j = \delta_{i,j}. \tag{3.3}$$

Equations (3.2) and (3.3) characterize the functionals y_i^* completely. Let $z_k^* = (R^*_{\mu_0})^{n+1-k} x_0^*$, $1 \leq k \leq n$.

LEMMA 3.1. *If $Ex = 0$, then $z_k^* x = 0$.*

Proof. If $Ex = 0$, then $(\mu - T - S)^{-1} x$ is regular near μ_0, or: the function $\psi(\mu)$ does not have a pole. This can happen only if the numerator of ψ has a zero of order n. Now, the numerator is $(x_0^* R_\mu x) R_\mu x_0$.

The vector $R_{\mu_0} x_0$ is never zero, hence the function $x_0^* R_\mu x$ has a zero of order n:

$$x_0^* R_{\mu_0} x = x_0^* R^2{}_{\mu_0} x = \cdots = x_0^* R^n{}_\mu x_0 = 0.$$

THEOREM 3.1. *The functionals y_j^* are given by $y_j^* = \sum_{k=1}^{n} \alpha_{j,k} z_k^*$, where the matrix $(\alpha_{j,k})$ is the inverse to the matrix $(z_j^*(y_k))$.*

Proof. The matrix $(z_j^*(y_k))$ is triangular and hence possesses an inverse. Let us define

$$y_j^* = \sum_{k=1}^{n} \alpha_{j,k} z_k^*,$$

where $\alpha_{j,k}$ are the elements of the inverse matrix. By Lemma 3.1 equation (3.2) is satisfied. Now equation (3.3) follows from

$$y_i^*(y_j) = \sum_{k=1}^{n} \alpha_{i,k} z_k^*(y_j) = \delta_{i,j}.$$

4. Spectral behavior of $T + S$ at isolated point of σ. Let μ_0 be an isolated point of σ. Thus

$$T = (\mu_0 - M) F + K(I - F), \tag{4.1}$$

where F is a projection, M a quasi nilpotent commuting with F, and $\mu_0 \notin \sigma$ (K restricted to $(I - F)X$). We shall start by some simple cases.

Case 1. Let $M = 0$. The spectrum of $T + S$ near μ_0 is the solution of

$$\begin{aligned} 1 = x_0^* R_\mu x_0 &= (x_0^* F x_0 / (\mu - \mu_0)) + x_0^* (\mu - K)^{-1} (I - F) x_0 \\ &= (x_0^* F x_0 / (\mu - \mu_0)) + f(\mu), \end{aligned} \tag{4.2}$$

where $f(\mu)$ is analytical near μ_0.

1. Let $x_0^* F x_0 \neq 0$.

In this case $x_0{}^*R_\mu x_0 \to \infty$ as $\mu \to \mu_0$, and there are only a finite number of solutions to equation (4.2). Now

$$\psi(\mu;x) = (-x_0{}^*Fx_0/(\mu-\mu_0) - f(\mu) + 1)^{-1}$$
$$(x_0{}^*Fx/(\mu-\mu_0) + x_0{}^*(\mu-K)^{-1}(1-F)x)$$
$$(Fx_0/(\mu-\mu_0) + (\mu-K)^{-1}(I-F)x_0).$$

The pole of $\psi(\mu)$ is simple with residue $-x_0{}^*Fx/x_0{}^*Fx_0$. Hence:

THEOREM 4.1. *If $M=0$ and $x_0{}^*Fx_0 \neq 0$, then the spectral projection E of $T+S$ at μ_0 is*

$$Ex = Fx - (x_0{}^*Fx)Fx_0/x_0{}^*Fx_0.$$

Thus the space of eigenvectors of $T+S$ with μ_0 as eigenvalue is given by $\{x \mid Fx = x$ and $x_0{}^*x = 0\}$. There is no nilpotent associated with $T+S$ at μ_0. Also $\mu_0 \notin \sigma(T+S)$ if and only if FX is a one dimensional space.

Proof. It was seen that $\psi(\mu)$ has a simple pole with residue

$$-(x_0{}^*Fx)Fx_0/x_0{}^*Fx_0.$$

Thus the projection of $T+S$ at μ_0 is $Fx - (x_0{}^*Fx)Fx_0/x_0{}^*Fx_0$; and there is no nilpotent. Now $\mu_0 \notin \sigma(T+S)$ if and only if $Fx = (x_0{}^*Fx)Fx_0/x_0{}^*Fx_0$.

If the above equation is satisfied then FX is one dimensional. Conversely, if FX is one dimensional, then the projection $(x_0{}^*Fx)Fx_0/x_0{}^*Fx_0$ is smaller than F and being both one-dimensional projections they are equal.

2. Let $x_0{}^*Fx_0 = 0$.

The equation for the spectrum will be:

$$x_0{}^*(\mu-K)^{-1}(I-F)x_0 = 1.$$

Thus in the neighbourhood of μ_0 there are only a finite number of point in $\sigma(T+S)$. In order to find the projection and nilpotent associated with $T+S$ at μ_0 we have to study the function

$$\psi(\mu)$$
$$= [1 - x_0{}^*(\mu-K)^{-1}(I-F)x_0]^{-1}(x_0{}^*Fx/(\mu-\mu_0) + x_0{}^*(\mu-K)^{-1}(I-F)x)$$
$$(Fx_0/(\mu-\mu_0) + (\mu-K)^{-1}(I-F)x_0).$$

2a. Let $\phi(\mu_0) \neq 0$; then

$$\psi(\mu) = \phi(\mu)^{-1}\{(\mu-\mu_0)^{-2}(x_0^*Fx)Fx_0 + (\mu-\mu_0)^{-1}[(x_0^*Fx)(\mu-K)^{-1}(I-F)x_0 + (x_0^*(\mu-K)^{-1}(I-F)x)Fx_0] + (x_0^*(\mu-K)^{-1}(I-F)x)(\mu-K)^{-1}(I-F)x_0\}.$$

But

$$\psi(\mu) = (\mu-T-S)^{-1} - (\mu-T)^{-1} = \sum_{i=0}^{\infty}(-N^iE)/(\mu-\mu_0)^{i+1} - F/(\mu-\mu_0) + (\text{terms analytical in } (\mu-\mu_0)),$$

where N and E are the quasi nilpotent and projection, respectively, associated with $T+S$ at μ_0. Comparing powers of $(\mu-\mu_0)^{-i+1}$ one gets $N^2=0$,

$$Nx = -\phi(\mu_0)^{-1}(x_0^*Fx)Fx_0,$$

$$E-F = \phi(\mu_0)^{-1}[(x_0^*Fx)(\mu_0-K)^{-1}(I-F)x_0 + (x_0^*(\mu_0-K)^{-1}(I-F)x)Fx_0] + (d/d\mu)\phi(\mu)^{-1}|_{\mu=\mu_0}(x_0^*Fx)Fx_0.$$

But $(d/d\mu)\phi(\mu)^{-1}|_{\mu=\mu_0} = -\phi(\mu_0)^{-2}x_0^*(\mu_0-K)^{-2}(I-F)x_0$. Thus $Ex-Fx = A_1x+A_2x$, where

$$A_1x = \phi(\mu_0)^{-1}(x_0^*Fx)(\mu_0-K)^{-1}(I-F)x_0,$$

$$A_2x = [\phi(\mu_0)^{-1}x_0^*(\mu_0-K)^{-1}(I-F)x - \phi(\mu_0)^{-2}(x_0^*(\mu_0-K)^{-2}(I-F)x_0)(x_0^*Fx)]Fx_0.$$

Both A_1 and A_2 are one dimensional operators and

$$A_1^2 = A_2^2 = FA_1 = A_1A_2 = 0,$$

$$A_1F = A_1, \qquad FA_2 = A_2, \qquad A_2F = -A_2A_1.$$

Finally let us check when $\mu_0 \notin \sigma(T+S)$. In this case $N = E = 0$. $N=0$ if and only if either $Fx_0 = 0$ or $F^*x_0^* = 0$. If $Fx_0 = 0$, then $A_2 = 0$ and $A_1 = \phi(\mu_0)^{-1}(x_0^*Fx)(\mu_0-K)^{-1}x_0$. In this case $E = F + A_1 \neq 0$ because if $F = -A_1$, then $FA_1 = -F^2 = F = 0$. If $F^*x_0^* = 0$, then $A_1 = 0$,

$$A_2 = \phi(\mu_0)^{-1}(x_0^*(\mu_0-K)^{-1}(I-F)x)Fx_0.$$

Now if $E = F + A_2 = 0$, then $F = -A_2$ and $A_2F = -F^2 = -F = -A_2A_1 = 0$ which is a contradiction.

Thus in this case $\mu_0 \in \sigma(T+S)$.

2b. Let $\phi(\mu_0) = 0$. Again we have

$$(\mu-T-S)^{-1} - (\mu-T)^{-1} = \phi(\mu)^{-1}[(x_0^*Fx/(\mu-\mu_0) + x_0^*(\mu-K)^{-1}(I-F)x)(Fx_0/(\mu-\mu_0) - (\mu-K)^{-1}(I-F)x_0)]$$

and $\phi(\mu_0)=\phi'(\mu_0)=\cdots=\phi^{(n-1)}(\mu_0)=0\neq\phi^{(n)}(\mu_0)$. The function $\psi(\mu)$ has a pole of order $n+2$ and $(-N)^{n+1}=(x_0{}^*Fx)Fx_0/\phi^{(n)}(\mu_0)$. It is possible to compute $E-F$ by taking the coefficient of $[\mu-\mu_0]^{-1}$ in $\psi(\mu)$: It is a sum of $n+2$ one-dimensional operators.

Case II. Let $M\neq 0$ but $M^n=0$. The function $\psi(\mu;x)$ has a finite pole, at most, at $\mu=\mu_0$: The denominator can have a zero of finite order and the numerator has a pole of order n. The spectrum equation

$$1=x_0{}^*R_\mu x_0-\sum_{i=0}^{n} x_0{}^*(-M)^iFx_0/(\mu-\mu_0)^{i+1}+\text{analytical terms}$$

has only a finite number of solutions near μ_0. Now the coefficients of negative powers of $(\mu-\mu_0)$ in $\psi(\mu)$ are finite sums of one dimensional operators. Thus:

THEOREM 4.2. *Let μ_0 be an isolated point of σ with F and M as the projection and nilpotent associated with T at μ_0. Let $M^n=0$. If E and N are the projection and nilpotent associated with $T+S$ at μ_0, then $E-F$ and $N-M$ are finite dimensional operators. Also $N^m=0$ for some integer m.*

If, in equation (4.1), M is a quasi nilpotent, the situation becomes more complicated. We shall give an example to show that one cannot expect an analog to Theorem 4.2.

Example. Let $X=C(01)$ and T be defined by $g=Tf$: $g(x)=\int_0^x f(t)dt$. Then $T^nf=\int_0^x[(x-t)^n/n!]f(t)\,dt$ and $R_\mu f=\mu^{-1}\int_0^x e^{(x-t)/\mu}f(t)\,dt$ for $\mu\neq 0$ Let $x_0(t)=1$ and $x_0{}^*f=f(1)$. The spectrum equation of $T+S$ is

$$1=x_0{}^*R_\mu x_0=\mu^{-1}\int_0^1 e^{(1-t)/\mu}\,dt=e^{1/\mu}-1,$$

or $\mu_k=[2\pi ik+\log 2]^{-1}$, $k=0,\pm 1,\pm 2,\cdots$. Now $x_0{}^*R_\mu{}^2x_0=-\mu^{-2}e^{1/\mu}$ $\neq 0$ and $[2\pi ik+\log 2]^{-1}$ is a simple eigenvalue. Also

$$E(\mu_k)f=\mu_k{}^2((1/\mu_k)\int_0^1 e^{(1-t)/\mu_k}f(t)\,dt)\,((1/\mu_k)\int_0^x e^{(1-t)/\mu_k}\,dt)$$

The eigenvectors of $T+S$ are

$$R_{\mu_k}x_0=\int_0^x(1/\mu_k)\,e^{(x-t)/\mu_k}\,dt=2e^{2\pi ikx}-1.$$

They generate a subspace of deficiency 1 but 0 is not an eigenvalue of $T+S$:

$$0=((T+S)f)(x)=\int_0^x f(t)\,dt+f(1)1.$$

Thus $((T+S)f)(0)=f(1)=0$ and $\int_0^x f(t)\,dt=0$. Therefore $f(t)=0$.

5. Examples: differential operators. We shall use our results to study differential operators on $C[a,b]$, $-\infty<a<b\leqq\infty$. Similar methods can be used in $L_p(a,b)$. Let Y be a Banach space of functions having n continuous derivatives, and let D be a continuous transformation from Y to X. $(X=C[a,b])$. Let $y_1^*,\cdots,y_n^*$ and $z_1^*,\cdots,z_n^*$ be functionals on Y. Define A as the restriction of D to Y_1, where

$$Y_1=\{f\,|\,f\in Y, y_i^*f=0, i=1,\cdots,n\}$$

and B the restriction of D to Y_2, where

$$Y_2=\{f\,|\,f\in Y, z_i^*f=0, i=1,\cdots,n\}.$$

Assumption. For some complex number λ, the operators $(A-\lambda)^{-1}$, $(B-\lambda)^{-1}$ are bounded everywhere defined operators on X. We may take $\lambda=0$. To avoid unnecessary complications, we assume that the functionals y_i^* (z_i^*) are independent. Let $z_1^*,\cdots,z_m^*$ be dependent on $y_1^*,\cdots,y_n^*$ but $z_{m+1}^*,\cdots,z_n^*$ not. The problem then is reduced to: $y_i^*=z_i^*$, $i=1,\cdots,m$, but $y_i^*,\cdots,y_m^*$, $y_{m+1}^*,\cdots,y_n^*$, $z_{m+1}^*,\cdots,z_n^*$ are linearly independent.

Let $Y_0=\{f\,|\,f\in Y$ and $y_i^*f=z_i^*f=0, i=1,\cdots,n\}$. Let $f_{m+1},\cdots,f_n$, $g_{m+1},\cdots,g_n$ be functions in Y such that

$$y_i^*(f_k)=0, \qquad y_i^*(g_j)=\delta_{i,j},$$
$$z_i^*(g_k)=0, \qquad z_i^*(f_j)=\delta_{i,j}.$$

There exist such functions because the functionals are independent. Then

$$Y_1=Y_0+\operatorname{span}\{f_{m+1},\cdots,f_n\}, \qquad Y_2=Y_0+\operatorname{span}\{g_{m+1},\cdots,g_n\}.$$

Let $\phi\in X$, $\phi=A\psi$, where $\psi\in Y_1$. Then $\psi=y_0+\sum_{i=m+1}^{n}(z_i^*\psi)f_i$, $y_0\in Y_0$, and $\phi=Ay_0+\sum_{i=m+1}^{n}(z_i^*\psi)Af_i=Dy_0+\sum_{i=m+1}^{n}(z_i^*\psi)Df_i$. Thus

$$(B^{-1}-A^{-1})\phi=(B^{-1}-A^{-1})Dy_0+\sum_{i=m+1}^{n}(z_i^*\psi)(B^{-1}-A^{-1})Df_i.$$

But on Y_0, $D=A=B$; hence $(B^{-1}-A^{-1})Dy_0=0$. Thus

$$(B^{-1}-A^{-1})\phi=\sum_{i=m+1}^{n}(z_i{}^*A^{-1}\phi)(B^{-1}-A^{-1})Df_i.$$

Now $f_i\in Y_1$, hence $Df_i=Af_i$. Let $B^{-1}Df_i=h_i$. Then $h_i\in Y_2$ and $Df_i=Bh_i=Dh_i$, or $D(h_i-f_i)=0$. Hence $B^{-1}Df_i=f_i+r_i$, where

$$Dr_i=0,\quad z_j{}^*(r_i)=-z_j{}^*(f_i)=-\delta_{i,j},\qquad i\geqq m, j\geqq 1, \tag{5.1}$$

and

$$(B^{-1}-A^{-1})\phi=\sum_{i=m+1}^{n}(z_i{}^*A^{-1}\phi)r_i. \tag{5.2}$$

Thus B^{-1} is an $n-m$ dimensional perturbation of A^{-1}.

Let, now, $b<\infty$ and let D be an n order differential operator. Choose $y_i{}^*f=f^{(i-1)}(a)$. Then A^{-1} is a quasi nilpotent operator. Let $z_i{}^*$ be arbitrary functionals on the space $C^n[a,b]$. Then all spectral properties of B can be studied by our method if there is a λ such that $(B-\lambda)^{-1}$ exists.

Example. Let D be a second order differential operator on $C(0,1)$. Let $y_1{}^*f=f(0)=z_1{}^*f$, $y_2{}^*f=f'(0)$, and $z_2{}^*$ be any functional on $C^2(0,1)$ independent of $y_1{}^*$ and $y_2{}^*$.

LEMMA. *The operator B^{-1} exists if and only if there exists a function $x_0(t)$ such that*

$$Dx_0=0,\qquad x_0(0)=0,\qquad z_2{}^*x_0=-1. \tag{5.3}$$

Proof. It was shown above that if B^{-1} exists there is such a function, namely $B^{-1}Df_1-f_1$, where

$$f_1(0)=f_1'(0)=0\quad\text{and}\quad z_2{}^*f_1=1.$$

Conversely if there exists such an x_0, then on Y_2, D is one to one: If $g\in Y_2$ and $Dg=0$, then $g=\alpha_1g_1+\alpha_2g_2$, where $Dg_1=Dg_2=0$, $g_1(0)=1$, $g_1'(0)=0$, $g_2(0)=0$, $g_2'(0)=1$. Now $g(0)=0$, hence $\alpha_1=0$, $g=\alpha_2g_2$. But x_0 satisfies the same conditions, hence $g=\alpha x_0$. Now $g\in Y_2$: $0=z_2{}^*g=z_2{}^*x_0=-\alpha$. Thus $g=0$. It remains to show that D is onto $C(0,1)$. Let $g_1\in Y$ and $g_1(0)=z_2{}^*(g_1)=0$, $g_1'(0)=1$.

Such a function exists if $y_2{}^*$ is independent of $y_1{}^*$ and $z_2{}^*$. Now A is onto, hence $Dg_1=Df+aDf_1$, where $y_1{}^*f=y_2{}^*f=z_2{}^*f=0$ and $y_1{}^*f_1=y_2{}^*f_1=0$, $z_2{}^*f_1=1$. If $a=0$, then $D(g_1-f)=0$ and $g_1-f\in Y_2$; but D is one to one on Y_2. Thus $Df_1=a^{-1}(Dg_1-Df)\in DY_2$ and $DY_2\supset DY_0+\{Df_1\}=DY_1$.

From the general discussion it follows that if x_0 exists and satisfies (5.3), then for every $x \in X$, $B^{-1}x = A^{-1}x + (z_2{}^*A^{-1}x)x_0$. Thus the spectrum of B^{-1} consists of the solutions of

$$1 = z_2{}^*A^{-1}(\mu - A^{-1})^{-1}x_0 = z_2{}^*(A\mu - I)^{-1}x_0 = \mu^{-1}z_2{}^*(A - 1/\mu)^{-1}x_0.$$

The operator B will have a multiple eigenvalue if

$$z_2{}^*(A\mu - 1)^{-1}x_0 = 1 \quad \text{and} \quad z_2{}^*A^{-1}(\mu - A^{-1})^{-2}x_0 = 0.$$

The second equation is equivalent to

$$0 = z_2{}^*A^{-1}(\mu^{-1}A)^2(A - 1/\mu)^{-2}x_0 = 1/\mu z_2{}^*(A - 1/\mu)^{-1}x_0$$
$$+ 1/\mu^2 z_2{}^*(A - 1\mu)^{-2}x_0 = 1 + 1/\mu^2 z_2{}^*(A - 1/\mu)^{-2}x_0.$$

The eigenvector for a given eigenvalue is

$$(\mu - A^{-1})^{-1}x_0 = (\mu A)(A - 1/\mu)^{-1}x_0.$$

University of California,
Berkeley, California.

REFERENCES.

[1] N. Dunford, "Spectral theory I, Convergence to projections," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 185-217.

[2] F. Wolf, "Perturbation by changes one-dimensional boundary conditions," *Indagationes Mathematicae*, vol. 18, No. 3 (1956), pp. 360-366.

WHITEHEAD PRODUCTS AND POSTNIKOV SYSTEMS.*

By Jean-Pierre Meyer.

0. Introduction. The Postnikov system of an arcwise-connected topological space X determines all homotopy operations of X, and, in particular, the Whitehead product pairing of the homotopy groups of X. The purpose of this note is to give an explicit description of this determination. In the process, some simple properties of Postnikov systems are obtained, in particular, a connection between the Postnikov system of X, J. H. C. Whitehead's "certain exact sequence" and the Hurewicz homomorphism for X.

In the last section, two applications are considered. Results of a similar nature have been obtained by M. G. Barratt (unpublished) and A. H. Copeland, Jr. [4]. Part of this work was supported by the Office of Naval Research, during the tenure by the author of a Research Associateship at Brown University. The author wishes to thank W. S. Massey for many helpful discussions.

1. The Hurewicz homomorphism. All spaces considered in this note are assumed to be simply-connected and to belong to the category, $\mathcal{W}$, of spaces having the homotopy-type of a CW-complex [8]. The reason for the latter assumption is that the composition of the two functors, singular complex, geometric realization, applied to an element of $\mathcal{W}$ gives another element of the same homotopy-type [8].

Consider a Postnikov system of a space X. We recall that this is a sequence of spaces and maps (X_n, p_n, q_n), $q_n\colon X\to X_n$, $p_n\colon X_n\to X_{n-1}$, such that:

(i) p_n is a fiber map with $K(\pi_n(X), n)$[1] as fiber.

(ii) q_n is an n-equivalence.

(iii) X_0 is a point.

(iv) $p_n q_n = q_{n-1}$.

For the construction and properties of such a Postnikov system, we refer to [5], [9]. The fiber of the composite map $p_m \cdots p_{n-1}p_n\colon X_n\to X_{m-1}$ $(m\leq n)$

* Received March 19, 1959.

[1] $K(\pi, n)$ denotes as usual an Eilenberg-MacLane space with $\pi_i = 0$, $i\neq n$, $\pi_n = \pi$.

is denoted by $X_{m,n}$. It is easily seen that the inclusion $X_{m,n} \to X_n$ induces isomorphisms of homotopy groups in dimensions m through n, so that $X_{m,n}$ has the same homotopy properties as X in that range of dimensions. The fiber $X_{n,n}$ will also be denoted by F_n and the inclusion $F_n \to X_n$ by i_n. We may assume that X_n is the fibre space induced by a map $k^{n+1}\colon X_{n-1} \to K(\pi_n(X), n+1)$ from the standard Serre path-space fibration

$$E(K(\pi_n(X), n+1)) \to K(\pi_n(X), n+1),^2$$

and therefore that F_n operates on X_n in a fiber-preserving manner. Each fibration $p_n\colon X_n \to X_{n-1}$ determines as usual a *Postnikov invariant* $k^{n+1} \in H^{n+1}(X_{n-1}; \pi_n)$, where $\pi_n = \pi_n(X)$.

(1.1) THEOREM. *$k^{n+1} = 0$ if and only if $i_{n*}\colon H_n(F_n) \to H_n(X_n)$ is a monomorphism onto a direct summand.*

Proof. As X is simply-connected, so is every X_n. The Serre exact sequence of $p_n\colon X_n \to X_{n-1}$, [12], is:

$$H_{n+1}(F_n) \xrightarrow[i_{n*}]{} H_{n+1}(X_n) \xrightarrow[p_{n*}]{} H_{n+1}(X_{n-1}) \xrightarrow[\tau]{} H_n(F_n) \to \cdots.$$

Since $H_{n+1}(F_n) = H_{n-1}(F_n) = 0$, we may write:

$$(1.2)_n \quad 0 \to H_{n+1}(X_n) \xrightarrow[p_{n*}]{} H_{n+1}(X_{n-1}) \xrightarrow[\tau]{} H_n(F_n) \xrightarrow[i_{n*}]{} H_n(X_n) \xrightarrow[p_{n*}]{} H_n(X_{n-1}) \to 0.$$

If i_{n*} is a monomorphism onto a direct summand, the exact sequence splits:

$$0 \to H_n(F_n) \to H_n(X_n) \to H_n(X_{n-1}) \to 0.$$

Hence, applying the functor $\mathrm{Hom}(\cdot\,; \pi_n)$, we obtain the split exact sequence:

$$0 \to \mathrm{Hom}(H_n(X_{n-1}); \pi_n) \to \mathrm{Hom}(H_n(X_n); \pi_n) \to \mathrm{Hom}(H_n(F_n); \pi_n) \to 0.$$

It follows from the universal coefficient theorem and the fact that

$$H^n(F_n; \pi_n) \approx \mathrm{Hom}(H_n(F_n); \pi_n) \text{ that } i_n^*\colon H^n(X_n; \pi_n) \to H^n(F_n; \pi_n)$$

is onto. Consider now the Serre cohomology sequence of p_n, which contains:

$$H^n(X_n; \pi_n) \xrightarrow[i_{n*}]{} H^n(F_n; \pi_n) \xrightarrow[\tau^*]{} H^{n+1}(X_{n-1}; \pi_n).$$

As i_n^* is onto, $k^{n+1} = -\tau^*(b^n) = 0$, where b^n is the basic cohomology class.

---

[2] $E(Y) \to Y$ denotes the Serre path-space fibration $p\colon E_{y_0,Y}(Y) \to Y$, $p(\omega) = \omega(1)$, $y_0 \in Y$.

The converse is trivial: if $k^{n+1}=0$, then $X_n \equiv X_{n-1} \times F_n$ and the projection $r: X_n \to F_n$ satisfies $ri_n =$ identity. Hence i_{n*} is a monomorphism onto a direct summand.

(1.3) COROLLARY. *$k^{n+1}=0$ if and only if the Hurewicz homomorphism $\mathcal{H}: \pi_n(X) \to H_n(X)$ is a monomorphism onto a direct summand.*

Proof. The diagram below is commutative.

$$\begin{array}{ccccc} H_n(F_n) & \xrightarrow{i_{n*}} & H_n(X_n) & \underset{\approx}{\xleftarrow{q_{n*}}} & H_n(X) \\ \approx \Big\uparrow \mathcal{H} & & \Big\uparrow \mathcal{H} & & \Big\uparrow \mathcal{H} \\ \pi_n(F_n) & \underset{\approx}{\xrightarrow{i_{n*}}} & \pi_n(X_n) & \underset{\approx}{\xleftarrow{q_{n*}}} & \pi_n(X). \end{array}$$

Since the indicated homomorphisms are isomorphisms, i_{n*} and $\mathcal{H}$ are equivalent, and the Corollary follows from (1.1).

(1.4) COROLLARY. *If X is an N-dimensional complex, then $k^{n+1}=0$ if and only if $\pi_n(X)=0$ $(n>N)$.*

Proof. If $n>N$, then $H_n(X)=0$, and $\mathcal{H}_n: \pi_n(X) \to H_n(X)$ is a monomorphism if and only if $\pi_n(X)=0$.

(1.5) PROPOSITION. *The following sequence is exact:*

$$\cdots H_{n+1}(X_{n+1}) \xrightarrow[(p_n p_{n+1})_*]{} H_{n+1}(X_{n-1}) \xrightarrow[\tau_n]{} H_n(F_n)$$
$$\xrightarrow[i_{n*}]{} H_n(X_n) \xrightarrow[(p_{n-1}p_n)_*]{} H_n(X_{n-2}) \to \cdots.$$

Proof. This follows immediately from the consideration of the exact sequences $(1.2)_{n+1}$, $(1.2)_n$, and $(1.2)_{n-1}$.

Although we shall not give the proof, this sequence is equivalent to J. H. C. Whitehead's sequence [16] and yields the isomorphism $\Gamma_n(X) \approx H_{n+1}(X_{n-1})$.

2. Whitehead products. Let $p, q > 1$, $s \in S^p$, $s' \in S^q$, and $e_p \in H_p(S^p, s)$, $e_q \in H_q(S^q, s')$ be generators. Then $H_i(S^p \times S^q, S^p \vee s^q) = 0$, $0 < i < p+q$, $H_{p+q}(S \ \times S^q, S^p \vee S^q) \approx Z$ and is generated by the cross-product, $e_p \times e_q$[1]. Here, as usual, $S^p \vee S^q$ denotes the subspace $(S^p \times s') \cup (s \times S^q)$. Let $\mathcal{H}_n$ denote the Hurewicz homomorphism in dimension n, and $i_p = \mathcal{H}_p(e_p)$, $i_q = \mathcal{H}_q(e_q)$; then $\pi_{p+q}(S^p \times S^q, S^p \vee S^q) \approx Z$, is generated by $\mathcal{H}_{p+q}(e_p \times e_q) = i_p \star i_q$ and $\partial(i_p \star i_q) \in \pi_{p+q-1}(S^p \vee S^p)$ is the Whitehead product $[i_{1*}i_p, i_{2*}i_q]$,

where $i_1\colon S^p\subset S^p\vee S^q$, $i_2\colon S^q\subset S^p\vee S^q$. If $\alpha\in\pi_p(X)$, $\beta\in\pi_q(X)$ are represented by $f\colon S^p\to X$, $g\colon S^q\to X$, then the Whitehead product $[\alpha,\beta]\in\pi_{p+q-1}(X)$ may be defined by $[\alpha,\beta]=(f\vee g)_*[i_1\cdot i_p, i_2\cdot i_q]$, where $f\vee g\colon S^p\vee S^q\to X$ is the usual "wedge" of the maps f, g, composed with the folding map $X\vee X\to X$.

(2.1) PROPOSITION. $\mathfrak{H}_{p+q-1}[\alpha,\beta]=0$.

Proof.
$$\begin{aligned}\mathfrak{H}_{p+q-1}[\alpha,\beta]&=\mathfrak{H}_{p+q-1}(f\vee g)_*[i_1\cdot i_p, i_2\cdot i_q]\\&=(f\vee g)_*\mathfrak{H}_{p+q-1}[i_1\cdot i_p, i_2\cdot i_q]=0,\end{aligned}$$
since $H_{p+q-1}(S^p\vee S^q)=0$.

(2.2) COROLLARY. *Identifying $H_{p+q-1}(F_{p+q-1})$ and $\pi_{p+q-1}(X)$ by the isomorphism exhibited in (1.3), there exists an element $\gamma\in H_{p+q}(X_{p+q-2})$ such that $\tau(\gamma)=[\alpha,\beta]$, where $\tau\colon H_{p+q}(X_{p+q-2})\to H_{p+q-1}(F_{p+q-1})$ is the transgression.*

Proof. This follows from (2.1), (1.3), and (1.4).

(2.3) Analogues of (2.1) and (2.2) hold for many homotopy operations.

In order to complete the connection between Postnikov systems and Whitehead products, it remains to identify an element $\gamma\in H_{p+q}(X_{p+q-2})$ satisfying (2.2). This will be done in the next section.

3. A homology class γ.

Assume X is $(p-1)$-connected.

(3.1) THEOREM. *In the fibration, $X_{p+q-2}\xrightarrow{P}X_{q-1}$ $(p\leqq q)$, the fiber $X_{q,p+q-2}$ has the homotopy type of a loop space ΩY, where Y is uniquely determined up to homotopy type. There exists a map (unique up to homotopy) $m\colon X_{q,p+q-2}\times X_{p+q-2}\to X_{p+q-2}$ such that*

a) *$m\mid X_{q,p+q-2}\vee X_{p+q-2}\simeq l\vee 1$, where l is the inclusion.*

b) *The following diagram is commutative:*

$$\begin{array}{ccc}X_{q,p+q-2}\times X_{p+q-2} & \xrightarrow{m} & X_{p+q-2}\\ \Big\downarrow{\scriptstyle *\times P} & & \Big\downarrow{\scriptstyle P}\\ *\times X_{q-1} & \xrightarrow{1} & X_{q-1}\end{array}$$

(*$*=$ single point, or the map into $*$, $1(*,x)=x$, $x\in X_{q-1}$*).

The proof will be given in the next section. It follows that one may

define a "generalized Pontrjagin product," [2], denoted as usual by $*$, $H_i(X_{q,p+q-2}) \otimes H_j(X_{p+q-2}) \to H_{i+j}(X_{p+q-2})$.

Let $\alpha \in \pi_p(X)$, $\beta \in \pi_q(X)$, $p \leqq q$. As follows from § 1, in computing $[\alpha, \beta]$, it suffices to consider $X_{p,p+q-1}$. We may therefore assume that $\pi_i(X) = 0$, $i < p$, $i > p+q-1$, so that $q_{p+q-1}\colon X \to X_{p+q-1}$ is a homotopy-equivalence under which we identify X and X_{p+q-1}. Let $f\colon S^p \to X$, $g\colon S^q \to X$ represent α, β, respectively, and $f'\colon S^p \to X_{p+q-2}$, $g'\colon S^q \to X_{p+q-2}$ be given by $f' = q_{p+q-2}f$, $g' = q_{p+q-2}g$. As the inclusion $l\colon X_{q,p+q-2} \to X_{p+q-2}$ induces isomorphisms of homotopy groups in dimensions q through $p+q-2$, we may factor g' through l up to a homotopy, i. e., there is a map $g''\colon S^q \to X_{q,p+q-2}$ such that $lg'' \simeq g'$. Let $\alpha' = \{f'\}$, $\beta' = \{g'\}$, $\beta'' = \{g''\}$.

(3.3) THEOREM. $\gamma = (-1)^{pq}\mathfrak{H}_q(\beta'') * \mathfrak{H}_p(\alpha')$ *satisfies* (2.2).

Proof. The terminology and some of the simpler results of [10] will be used.

a) Let $W\colon S^{p+q-1} \to S^p \vee S^q$ represent $[i_1 \circ i_p, i_2 \circ i_q]$. It is easily seen from § 2 and is well-known [13] that $S^p \times S^q$ results from attaching a $(p+q)$-cell to $S^p \vee S^q$ by W, i. e., $S^p \times S^q = (S^p \vee S^q) \cup_W CS^{p+q-1}$. Hence ([10], § 5),

$$S^{p+q-1} \xrightarrow{W} S^p \vee S^q \xrightarrow{j} S^p \times S^q \xrightarrow{\eta} S^{p+q} \to \cdots$$

is a coexact sequence of spaces; j is the inclusion, and η is the identification map shrinking $S^p \vee S^q$ to a point.

b) Let $\pi = \pi_{p+q-1}(X)$; it follows from § 1 and a slight extension of [10], Lemma 2. 1, that

$$\cdots \to K(\pi, p+q-1) \xrightarrow{i} X_{p+q-1} \xrightarrow{p} X_{p+q-2} \xrightarrow{k} K(\pi, p+q).$$

is an exact sequence of spaces; $i = i_{p+q-1}$, $p = p_{p+q-1}$, $k = k^{p+q}$.

Consequently, all rows and columns of the following diagram are exact:

$$\begin{tikzcd}
 & & \pi(S^p \vee S^q; K(\pi, p+q-1)) \arrow[r, "W^\#"] \arrow[d, "i_\#"] & \pi(S^{p+q-1}; K(\pi, p+q-1)) \arrow[d, "i_\#", "\approx"'] \\
 & & \pi(S^p \vee S^q; X_{p+q-1}) \arrow[r, "W^\#"] \arrow[d, "p_\#", "\approx"'] & \pi(S^{p+q-1}; X_{p+q-1}) \arrow[d, "p_\#"] \\
\cdots \to \pi(S^{p+q}; X_{p+q-2}) \arrow[r, "\eta^\#"] \arrow[d, "k_\#"] & \pi(S^p \times S^q; X_{p+q-2}) \arrow[r, "j^\#", "\approx"'] \arrow[d, "k_\#"] & \pi(S^p \vee S^q; X_{p+q-2}) \arrow[r, "W^\#"] \arrow[d, "k_\#"] & \pi(S^{p+q-1}; X_{p+q-2}) \arrow[d, "k_\#"] \\
\cdots \to \pi(S^{p+q}; K(\pi, p+q)) \arrow[r, "\eta^\#", "\approx"'] & \pi(S^p \times S^q; K(\pi, p+q)) \arrow[r, "j^\#"] & \pi(S^p \vee S^q; K(\pi, p+q)) \arrow[r, "W^\#"] & \pi(S^{p+q-1}; K(\pi, p+q)).
\end{tikzcd}$$

The set of homotopy classes of maps $A \to B$, $\pi(A;B)$ possesses a canonically defined group structure if A is a suspension, or B is a loop-space (see, for example, [10]); hence, in the above diagram, all sets are groups, with the possible exception of $\pi(S^p \times S^q; X_{p+q-2})$, and all maps are homomorphisms. The latter is also a group; in fact, $j^{\#}: \pi(S^p \times S^q; X_{p+q-2}) \to \pi(S^p \vee S^q; X_{p+q-2})$ is a one-one correspondence (see (4.11)). To simplify the notation, we shall not distinguish between a map and its homotopy class. Let $a \vee b: S^p \vee S^q \to X_{p+q-2}$, and $b': S^q \to X_{q,p+q-2}$ be such that $lb' \simeq b$. Then the map $S^p \times S^q \xrightarrow{\xi} S^q \times S^p \xrightarrow{b' \times a} X_{q,p+q-2} \times X_{p+q-2} \xrightarrow{m} X_{p+q-2}$, where ξ is the homeomorphism interchanging the factors, has the property that

$$m(b' \times a)\xi j \simeq a \vee b$$

and therefore $j^{\#}(m(b' \times a)\xi) = a \vee b$.

Recalling that $\pi(A; K(\pi, n)) \approx H^n(A; \pi)$, the following groups vanish:

$$\pi(S^p \vee S^q; K(\pi, p+q-1)), \quad \pi(S^{p+q-1}; X_{p+q-2}), \quad \pi(S^{p+q}; X_{p+q-2}),$$
$$\pi(S^p \vee S^q; K(\pi, p+q)), \text{ and } \pi(S^{p+q-1}; K(\pi, p+q));$$

so the indicated homomorphisms are isomorphisms. Following the usual method for defining functional operations, we obtain the homomorphisms:

$$i_{\#}^{-1} W^{\#} p_{\#}^{-1}: \pi(S^p \vee S^q; X_{p+q-2}) \to \pi(S^{p+q-1}; K(\pi, p+q-1))$$
$$\eta^{\#-1} k_{\#} j^{\#-1}: \pi(S^p \vee S^q; X_{p+q-2}) \to \pi(S^{p+q}; K(\pi, p+q)).$$

By Theorem 7.1 of [10], or a straightforward direct verification, one sees that

$$\phi \eta^{\#-1} k_{\#} j^{\#-1} = i_{\#} W^{\#} p_{\#}^{-1},$$

where ϕ is the classical Hurewicz isomorphism

$$\pi(S^{p+q}; K(\pi, p+q)) \to \pi(S^{p+q-1}; K(\pi, p+q-1)).$$

Identifying $\pi(S^{p+q}; K(\pi, p+q))$, $\pi(S^{p+q-1}; K(\pi, p+q-1))$, $\pi(S^{p+q-1}; X_{p+q-1})$ under ϕ and $i_{\#}$, we obtain $\eta^{\#-1} k_{\#} j^{\#-1} = W^{\#} p_{\#}^{-1}$.

Consider $f' \vee g' \in \pi(S^p \vee S^q; X_{p+q-2})$. Then $p_{\#}^{-1}(f' \vee g') = f \vee g$, and $W^{\#} p_{\#}^{-1}(f' \vee g') = (f \vee g)_* W = (f \vee g)_*[i_1 \cdot i_p, i_2 \cdot i_q] = [\alpha, \beta]$. On the other hand, $j^{\#-1}(f' \vee g') = m(g'' \times f')\xi$, and $k_{\#} j^{\#-1}(f' \vee g') = km(g'' \times f')\xi$.

$$\begin{tikzcd}
\pi(S^{p+q}; K(\pi, p+q)) \arrow[rr, "\eta^{\#}"] \arrow[dr, "\theta'"'] & & \pi(S^p \times S^q; K(\pi, p+q)) \arrow[dl, "\theta"] \\
& \pi &
\end{tikzcd}$$

where $\theta'(a) = \langle a^* b^{p+q}, s^{p+q}\rangle$, and $\theta(b) = \langle b^* b^{p+q}, s^p \times s^q\rangle$; here $s^i \in H_i(S^i)$ is a generator and $\langle \cdot, \cdot \rangle$ denotes the Kronecker pairing of cohomology and homology, and θ, θ' are isomorphisms. Identifying under these isomorphisms, one obtains

$$\begin{aligned}
\eta^{\#-1} k_\# j^{\#-1}(f' \vee g') &= \theta k_\# j^{\#-1}(f' \vee g') \\
&= \theta k m (g'' \times f')\xi \\
&= \langle \xi^* (g'' \times f')^* m^* k^* b^{p+q}, s^p \times s^q\rangle \\
&= (-1)^{pq} \langle (g'' \times f')^* m^* k^{p+q}, s^p \times s^q\rangle \\
&= (-1)^{pq} \langle k^{p+q}, m_* (g''_* s^q \times f'_* s^p)\rangle \\
&= (-1)^{pq} \langle k^{p+q}, m_* (\mathcal{H}_q(\beta'') \times \mathcal{H}_p(\alpha'))\rangle \\
&= (-1)^{pq} \langle k^{p+q}, \mathcal{H}_q(\beta'') * \mathcal{H}_p(\alpha')\rangle
\end{aligned}$$

and, therefore,

$$\langle k^{p+q}, \mathcal{H}_q(\beta'') * \mathcal{H}_p(\alpha')\rangle = (-1)^{pq}[\alpha, \beta],$$

from which (3.3) follows.

4. Proof of (3.1). In the fibration $P\colon X_{p+q-2} \to X_{q-1}$, the fiber $X_{q,p+q-2}$ has non-vanishing homotopy groups in dimensions $q, q+1, \cdots, p+q-2$. Since $p+q-2 \leqq 2q-2$, the first statement of (3.1) follows from the well-known theorem to the effect that the Postnikov invariants of the loop-space of a space are the suspensions of the original Postnikov invariants. One merely constructs the space Y inductively, by "de-suspending" the Postnikov invariants of $X_{q,p+q-2}$. This can be done, at each stage, uniquely, by Prop. 10, Ch. IV of [12]. See, for example, [7] and [14].

Before we proceed to the proof of (3.1), note that, in two special cases of particular interest, (3.1) is now trivial:

(1) $p = q$; in which case X_{q-1} is a single point, so

$$X_{q,p+q-2} = X_{p+q-2}.$$

(2) $\pi_i(X) = 0$, $q < i < p+q-1$; in which case

$$X_{p+q-2} = X_q \text{ and } X_{q,p+q-2} = F_q = K(\pi_q, q).$$

(4.1) LEMMA. *If A is $(p-1)$-connected, B is $(q-1)$connected, and $\pi_i(C) = 0$, $i \geqq p+q-1$, then*

$$r^\#\colon \pi(A \times B; C) \to \pi(A \vee B; C)$$

is one-one, where $r\colon A \vee B \to A \times B$ is the inclusion.

Proof. Let $f \in \pi(A \vee B; C)$, then the successive obstructions to extending f lie in the groups $H^{i+1}(A \times B, A \vee B; \pi_i(C))$, which all vanish. On the other hand, let $g, h \in \pi(A \times B; C)$ and $r^{\#}(g) = r^{\#}(h)$. Then the obstructions to a homotopy between g and h lie in the groups $H^i(A \times B, A \vee B; \pi_i(C))$, which also vanish.

Applying (4.1) to $A = X_{p+q-2}$, $B = X_{q,p+q-2}$, $C = X_{p+q-2}$, one obtains a). To verify b), apply (4.1) successively to $A = X_\alpha$, $B = X_{q,\alpha}$, $C = X_\alpha$ or $X_{\alpha-1}$, $q-1 \leqq \alpha \leqq p+q-2$. In the following diagram,

$$\begin{tikzcd}
\pi(X_{q,\alpha} \times X_\alpha; X_\alpha) \arrow[r, "r^{\#}"] \arrow[d, "p_{\alpha\#}"] & \pi(X_{q,\alpha} \vee X_\alpha; X_\alpha) \arrow[d, "p_{\alpha\#}"] \\
\pi(X_{q,\alpha} \times X_\alpha; X_{\alpha-1}) \arrow[r, "r^{\#}"] & \pi(X_{q,\alpha} \vee X_\alpha; X_{\alpha-1}) \\
\pi(X_{q,\alpha-1} \times X_{\alpha-1}; X_{\alpha-1}) \arrow[r, "r^{\#}"] \arrow[u, "(t_\alpha \times p_\alpha)^{\#}"] & \pi(X_{q,\alpha-1} \vee X_{\alpha-1}; X_{\alpha-1}) \arrow[u, "(t_\alpha \vee p_\alpha)^{\#}"']
\end{tikzcd}$$

the maps $r^{\#}$ are one-one, and $p_\alpha: X_\alpha \to X_{\alpha-1}$, $t_\alpha: X_{q,\alpha} \to X_{q,\alpha-1}$ are the natural fiber maps. Let $s_\alpha: X_{q,\alpha} \to X_\alpha$ be the inclusion, and $m_\alpha: X_{q,\alpha} \times X_\alpha \to X_\alpha$ be such that $r^{\#}(m_\alpha) = s_\alpha \vee 1$. Then,

$$\begin{aligned}(t_\alpha \vee p_\alpha)^{\#}(s_{\alpha-1} \vee 1) &= t_\alpha^{\#} s_{\alpha-1} \vee p_\alpha^{\#} 1 = p_{\alpha\#} s_\alpha \vee p_{\alpha\#}(1) \\ &= p_{\alpha\#}(s_\alpha \vee 1).\end{aligned}$$

Hence,

$$(t_\alpha \times p_\alpha)^{\#} m_{\alpha-1} = p_{\alpha\#} m_\alpha, \text{ i.e., } p_\alpha m_\alpha = m_{\alpha-1}(t_\alpha \times p_\alpha),$$
$$\alpha = q, q+1, \cdots, p+q-2.$$

Letting $m = m_{p+q-2}$, and using the last identity repeatedly, one easily verifies b).

(4.2) *Remark.* A natural conjecture, which I have been unable to prove, is that a stronger form of (3.1) holds.[3] Namely: P is a principal fiber map in the sense that $P: X_{p+q-2} \to X_{q-1}$ is of the same fiber homotopy type as a fiber map induced from $E(Y) \to Y$ by a map $X_{p+q-2} \to Y$. This is particularly plausible in view of the following: let $F = X_{q,p+q-2}$. Then $F^F \equiv \Omega Y^{\Omega Y}$, and ΩY may be identified with a subspace (in fact, a direct summand) of $\Omega Y^{\Omega Y}$. Furthermore, the "twisting map" [3] of P, $\Omega X_{p+q-2} \to F^F$, composed with the homotopy equivalence $F^F \to \Omega Y^{\Omega Y}$ can be deformed into a map $\Omega X_{p+q-2} \to \Omega Y$ and this latter map can be "de-looped" to a map $X_{p+q-2} \to Y$.

[3] Added in proof: in the meantime, this has been verified; a proof will appear in a note, "On principal fibrations."

5. Applications.

(5.1) THEOREM. *Let X be a space such that $\pi_i(X) \neq 0$ only for $n \leqq i \leqq 2n-1$. Then X is an H-space (in fact, has the homotopy type of a loop-space), if and only if the Whitehead product $\pi_n(X) \otimes \pi_n(X) \to \pi_{2n-1}(X)$ vanishes.*

Proof. If X is an H-space, it is a classical result that all Whitehead products vanish. Conversely, let $[\alpha, \beta] = 0$, $\alpha, \beta \in \pi_n(X)$, and let m be the multiplication on X_{2n-2}. Then it follows from (3.3) that k^{2n} is a primitive element of $H^{2n}(X_{2n-2}; \pi_{2n-1}(X))$. As X_{2n-2} is a loop-space, it follows from Corollary 6.3 of [15] that k^{2n} is a suspension element, and therefore that X_{2n-1} is a loop-space. Since $q_{2n-1}: X \to X_{2n-1}$ is a homotopy-equivalence, the theorem is proved.

(5.2) THEOREM. *Let X be a space such that $\pi_n(X) \approx Z$ and $\pi_i(X) = 0$, $n < i < 2n-1$, n odd. Then the Whitehead product $\pi_n(X) \otimes \pi_n(X) \to \pi_{2n-1}(X)$ vanishes.*

Proof. One may assume, as seen in § 3, that $\pi_i(X) = 0$, $i < n$. Then $X_n = X_{n+1} = \cdots = X_{2n-2} = K(Z, n)$. Let $b \in H_n(Z, n)$ denote a generator; then the Pontrjagin product $b * b \in H_{2n}(Z, n)$ vanishes, [11]. Hence, the conclusion follows from (3.3).

Let L be a graded Lie ring in the sense of Hilton [6]. The following interesting problem was pointed out by N. Stein: under what conditions can L be realized as the homotopy ring of an arcwise-connected space X? The example (5.2) indicates that additional conditions are required.

The applicability of the foregoing theory to the computation of Whitehead products clearly depends on the computation of the "generalized Pontrjagin product." This product can be computed in certain simple cases, but no general method is known to the author. Such a method would be of considerable interest.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. L. Blakers and W. S. Massey, "Products in homotopy theory," *Annals of Mathematics*, vol. 58 (1953), pp. 295-324.

[2] R. Bott and H. Samelson, "On the Pontryagin product in spaces of paths," *Commentarii Mathematici Helvetici*, vol. 27 (1953), pp. 320-337.

[3] E. H. Brown, Jr., "Twisted Tensor Products I," *Annals of Mathematics*, vol. 69 (1959), pp. 223-246.

[4] A. H. Copeland, Jr., "On H-spaces with two non-trivial homotopy groups," *Proceedings of the American Mathematical Society*, vol. 8 (1957), pp. 184-191. Also, Abstract 271, *Bulletin of the American Mathematical Society*, vol. 62 (1956), p. 180.

[5] A. Heller, "Homotopy resolutions of semi-simplicial complexes,' *Transactions of the American Mathematical Society*, vol. 80 (1955), pp. 299-344.

[6] P. J. Hilton, "Note on quasi-Lie rings," *Fundamenta Mathematicae*, vol. 43 (1956), pp. 230-237.

[7] Y. Inoue, "A note on the singular homotopy type of spaces," *Proceedings of the Japan Academy*, vol. 33 (1957), pp. 90-91.

[8] J. W. Milnor, "On spaces having the homotopy-type of a CW-complex," *Transactions of the American Mathematical Society*, vol. 90 (1959), pp. 272-280.

[9] J. C. Moore, "Semi-simplicial complexes and Postnikov systems," *Proceedings of the International Symposium on Algebraic Topology*, Mexico, 1956.

[10] F. P. Peterson, "Functional cohomology operations," *Transactions of the American Mathematical Society*, vol. 86 (1957), pp. 197-211.

[11] H. Samelson, "Groups and spaces of loops," *Commentarii Mathematici Helvetici*, vol. 28 (1954), pp. 278-286.

[12] J.-P. Serre, "Homologie singulière des espaces fibrés,' *Annals of Mathematics*, vol. 54 (1951), pp. 425-505.

[13] N. E. Steenrod, "Cohomology invariants of mappings," *Annals of Mathematics*, vol. 50 (1949), pp. 954-988.

[14] H. Suzuki, "On the Eilenberg-MacLane invariants of loop spaces," *Journal of the Mathematical Society of Japan*, vol. 8 (1956), pp. 40-48.

[15] G. W. Whitehead, "On the homology suspension," *Annals of Mathematics*, vol. 62 (1955), pp. 254-268.

[16] J. H. C. Whitehead, "A certain exact sequence," *ibid.*, vol. 52 (1950), pp. 51-110.

A CHARACTERIZATION OF THE SIMPLE GROUPS $SL(2, 2^a)$.*

By WALTER FEIT.[1]

1. Introduction. Let G be a finite group of order g. For a prime p which divides g, let S_p denote a Sylow p-group G. In this paper we will study groups which satisfy the following condition.

(*) *The group G has even order, and for any element u of order two in S_2 the centralizer of u is contained in the centralizer of S_2.*

The groups $SL(2, 2^a)$, consisting of all two by two unimodular matrices with entries in the finite field of 2^a elements, satisfy condition (*). For $a > 1$, $SL(2, 2^a)$ is always a simple group. The main purpose of this paper is to prove that conversely, the groups $SL(2, 2^a)$ with $a > 1$, are the only simple groups which satisfy condition (*). The following more general result is proved.

THEOREM 1. *If the group G satisfies condition (*), then at least one of the following statements is true:*

(I) *Every Sylow 2-group of G is cyclic.*

(II) *The Sylow 2-group of G is a normal subgroup of G.*

(III) *The group G is the direct product of a group D of odd order and a group L which is isomorphic to $SL(2, 2^a)$ for some $a > 1$.*

It is well known that a Sylow 2-group of a simple group cannot be cyclic, hence Theorem 1 immediately yields the characterization of the groups $SL(2, 2^a)$ mentioned above.

COROLLARY. *Let G be a non-cyclic simple group. The group G satisfies condition (*) if and only if G is isomorphic to $SL(2, 2^a)$ for some $a > 1$.*

The following two results can easily be deduced from Theorem 1.

THEOREM 2. *Let G be a group of order $4g'$, where g' is odd. Assume that the intersection of the Sylow 2-group S_2 with any other Sylow 2-group of G has order one. Then S_2 is either cyclic or a normal subgroup of G, or G*

* Received April 1, 1959.

[1] National Science Foundation Fellow 1958-59.

is the direct product of a group D of odd order with a group L isomorphic to $SL(2,4)$. In particular, if G is a simple group, then G is isomorphic to $SL(2,4)$, the simple group of order 60.

THEOREM 3. *Suppose that G is a non-cyclic simple group of even order with abelian Sylow 2-groups. Furthermore, assume that for every element u of order two, the centralizer of u is nilpotent. Then G is isomorphic to $SL(2,2^a)$ for some $a>1$.*

As a special case of Theorem 3, one gets a result recently proved by M. Suzuki [6], which states that the groups $SL(2,2^a)$ with $a>1$ are the only non-cyclic simple groups of even order in which the centralizer of every element of order two is abelian. This result contains earlier results of R. Brauer, M. Suzuki and G. E. Wall [2] as special cases. The assumption that the Sylow 2-groups of G are abelian is essential in Theorem 3, since for example, the simple group $PSL(2,7)$ of order 168 has the property that the centralizer of every element of order two is nilpotent. However $PSL(2,7)$ is not isomorphic to $SL(2,2^a)$ for any value of a.

After some preliminary results have been proved in Section 3, the proof of Theorem 1 is broken up into two cases. The first case is handled in Section 4 and depends on a theorem proved in a previous paper [5]. The second case, which is handled in Sections 5 and 6, depends on some arithmetical properties of characters which are developed in Section 6.

2. Notation and prerequisites. For any subset T of a group G, $C(T)$, $N(T)$, $|T|$ will mean respectively the centralizer, normalizer, and number of elements in T. For a cyclic subgroup $T=\{x\}$ of G, $C^*(T)=C^*(x)$ will denote the subgroup of G consisting of all elements y in G with $yxy^{-1}=x$ or x^{-1}. It is easily seen that $[C^*(x):C(x)]\leqq 2$. An element of order two is called an involution. An element x which is conjugate to its inverse is said to be real.[2] It is easily seen that if an element x is real, then either $x^2=1$ or $[C^*(x):C(x)]=2$.

For any complex valued functions ξ_1, ξ_2 on G, the hermitian product $(\xi_1,\xi_2)_G$ is defined by

$$(\xi_1,\xi_2)_G=1/g\ \sum_G \xi_1(x)\overline{\xi_2(x)},$$

and the norm by

$$\|\xi_1\|^2_G=(\xi_1,\xi_1)_G.$$

The subscript G will be dropped in cases where it is clear from the context which group is involved.

[2] See [1] for these definitions.

A group which contains a proper normal subgroup M with the property that no element in $M-\{1\}$ commutes with any element not in M, is called a Frobenius group. The subgroup M is called the regular subgroup. Elementary properties of such groups can be found in [4].

At the beginning of Sections 3, 4, 5, and 6 some hypotheses are stated. Each lemma in any of these sections is assumed to satisfy the hypotheses stated in the beginning of the section containing it.

Finally we state here without proof some well known properties of the group $SL(2, 2^a)$ which will be needed later.

LEMMA 2.1. *Let L be a group isomorphic to $SL(2,2^a)$ for some $a>1$, let S_2 be a Sylow 2-group of L. Then L is a simple group of order $(2^a-1)2^a(2^a+1)$ with the following properties.*

(i) *S_2 is the direct product of groups of order two. Any two elements in $S_2-\{1\}$ are conjugate in $N(S_2)$, $|N(S_2)|=2^a(2^a-1)$.*

(ii) *Any two distinct Sylow 2-groups of L generate L.*

(iii) *When considered as a permutation group on its Sylow 2-groups, L is triply transitive and no element leaves three Sylow 2-groups fixed.*

(iv) *If x is not in $N(S_2)$, then $N(xS_2x^{-1})\cap N(S_2)$ is a cyclic group of order (2^a-1).*

(v) *L contains a cyclic group K of order (2^a+1) and a cyclic group Q of order (2^a-1). Every element in L is real, $N(K)=C^*(K)$, $N(Q)=C^*(Q)$. Every element in L of order greater than two is conjugate to an element either in K or in Q.*

3. Preliminary remarks. If Theorem 1 is false, then there exists a group of minimum order which satisfies condition (*) but for which the conclusion of Theorem 1 does not hold. Throughout the remainder of the paper, the following will be assumed.

HYPOTHESIS. *The group G of order g satisfies condition (*) but not the conclusion of Theorem* 1. *Furthermore the conclusion of Theorem* 1 *holds for any group of order strictly smaller than g which satisfies condition* (*).

The remainder of the paper is devoted to deriving a contradiction from this Hypothesis.

It is easily seen that every proper subgroup of G of even order satisfies condition (*), and hence the conclusion of Theorem 1. Furthermore g is divisible by 4.

LEMMA 3.1. *A Sylow 2-group S_2 of G is abelian of order $2^a > 2$. If x is any element not in $N(S_2)$, then $S_2 \cap xS_2x^{-1} = \{1\}$. $C(S_2) = S_2 \times H$ for some group H of odd order. If S_2 contains q involutions, then $[N(S_2): C(S_2)] = q$ and any two involutions in S_2 are conjugate in $N(S_2)$. Furthermore, the only real elements which commute with an involution are involutions or the identity element.*

Proof. The first two statements are self evident. It follows from a theorem of Burnside ([3], p. 327) that $C(S_2) = S_2 \times H$.

Suppose $xS_2x^{-1} \neq S_2$, let u, v be involutions in S_2. It is immediate from condition (*) that no involution commutes with both v and xux^{-1}, hence by [1] Lemma 3A, v is conjugate to xux^{-1}, therefore v is conjugate to u. Any two involutions in S_2 are conjugate in G. If u, xux^{-1} are both in S_2, then

$$xux^{-1} \in S_2 \cap xS_2x^{-1},$$

therefore x must be in $N(S_2)$. Any two involutions in S_2 are conjugate in $N(S_2)$.

The group $N(S_2)$ acts as a transitive permutation group on the q involutions in S_2, hence $[N(S_2): C(u)] = q$ for any involution $u \in S_2$. By assumption $C(u) = C(S_2)$.

If x is real and $x \in C(u)$ for some involution $u \in S_2$, then $x \in C(S_2)$. Therefore $[G: C(x)]$ is odd, hence $C^*(x) = C(x)$. Consequently $x = x^{-1}$ as was to be shown.

LEMMA 3.2. *Every coset of $C(S_2)$, other than those in $N(S_2)$, contains exactly one involution.*

Proof. The number of involutions in G is $[G: C(S_2)]$, since by Lemma 3.1 any two involutions are conjugate in G, and by assumption $C(u) = C(S_2)$ for any involution $u \in S_2$. The group $N(S_2)$ contains q involutions and q cosets of $C(S_2)$, hence the lemma will be proved once it is shown that no coset of $C(S_2)$, other than $C(S_2)$, can contain more than one involution. Suppose u, v are distinct involutions and $uC(S_2) = vC(S_2)$, then $uv \in C(S_2)$. By Lemma 3.1, uv is an involution, hence $uv \in S_2$, and $u \in C(uv) = C(S_2)$. Therefore $uC(S_2) = vC(S_2) = C(S_2)$ as was to be shown.

LEMMA 3.3. *For $y \in N(S_2)$, either $C(y) \subset N(S_2)$ or y commutes with an involution not in S_2. If y is a real element of order greater than two in $N(S_2)$, then $C(y) \subset N(S_2)$.*

Proof. Suppose some element not in $N(S_2)$ commutes with y, then by Lemma 3.2, this element is of the form xu, where $x \in C(S_2)$, and u is an

involution not in S_2. Consequently $uyu = x^{-1}yx$, hence $y^{-1}uyu = y^{-1}x^{-1}yx$. It is easily seen that the commutator subgroup of $\{y\}C(S_2)$ is contained in $C(S_2)$. Since $y^{-1}uy$ is an involution, this implies that $y^{-1}uyu$ is a real element in $C(S_2)$. By Lemma 3.1, $y^{-1}uyu = 1$ or $y^{-1}uyu$ is an involution. In the latter case $y^{-1}uyu \in S_2$ and $u \in N(S_2)$, hence $u \in S_2$. This is impossible, therefore $y^{-1}uyu = 1$, hence $y \in C(u)$.

For a real element y of order greater than two, $C(y)$ cannot contain any involution by Lemma 3.1, hence $C(y) \subset N(S_2)$.

LEMMA 3.4. *$N(S_2)$ contains at least $(g/qc(S_2) - 1)(q-1)$ real elements, where $c(S_2) = |C(S_2)|$.*

Proof. Let u_1, u_2, v_1, v_2, be distinct involutions not in $N(S_2)$ such that

$$u_1N(S_2) = u_2N(S_2) \neq v_1N(S_2) = v_2N(S_2).$$

Suppose that $u_1u_2 = v_1v_2$. Then $y = u_1u_2$ is a real element in $N(S_2)$, and u_1, v_1 are in $C^*(y)$. Hence $u_1v_1 \in C(y)$. Therefore by Lemma 3.3, $u_1v_1 \in N(S_2)$ which is impossible by the choice of u_1 and v_1.

By Lemma 3.2 each coset of $N(S_2)$ distinct from $N(S_2)$ contains q involutions. For such a coset, let $u_1, \cdots, u_q$ be the involutions in $u_1N(S_2)$, then $u_1u_2, \cdots, u_1u_q$ are $q-1$ distinct real elements in $N(S_2)$. By what has just been proved, $u_1u_i \neq v_1v_j$ for any i, j where $v_1N(S_2) = v_jN(S_2)$ is a coset of $N(S_2)$ distinct from $u_1N(S_2)$. Therefore $N(S_2)$ contains at least $(g/qc(S_2) - 1)(q-1)$ distinct real elements.

LEMMA 3.5. *If G contains a subgroup L which is isomorphic to $SL(2, 2^a)$, then $N(L) = L \times C(L)$.*[3]

Proof. The group $N(L)$ is triply transitive when considered as a group of permutations on the Sylow 2-groups of L. Let C be the subgroup of $N(L)$ consisting of those elements which leave three Sylow 2-groups S_2, S'_2, S''_2 fixed, then $N(L) = LC$ and

$$C = N(S_2) \cap N(S'_2) \cap N(S''_2).$$

Let Q', Q'' be defined by

$$Q' = N(S_2) \cap N(S'_2) \cap L, \qquad Q'' = N(S_2) \cap N(S''_2) \cap L.$$

By Lemma 2.1, Q' and Q'' are cyclic groups of order $q = 2^a - 1$. If $x \in C$, then it is easily seen that $x \in N(Q') \cap N(Q'') \cap N(S_2)$. By Lemma 2.1,

$$C(S_2) \cap Q' = C(S_2) \cap Q'' = \{1\}.$$

---

[3] The arguments used in Lemmas 3.5, 3.6 and 3.7 are similar to those used in [6] for the proof of some analogous lemmas.

By Lemma 3.1, $[N(S_2):C(S_2)]=q$. Hence Q' and Q'' are isomorphic to their images in $N(S_2)/C(S_2)$. Therefore no two distinct elements of Q' or Q'' are conjugate in $N(S_2)$. Hence $x\in C(Q')\cap C(Q'')$. Therefore x commutes with every element in the group generated by Q' and Q''. Since Q', Q'' are both contained in $N(S_2)\cap L$, it is easily seen that they generate $N(S_2)\cap L$. Hence $x\in C(S_2)$. By a similar argument $x\in C(S'_2)$, therefore $x\in C(L)$ since S_2 and S'_2 generate L. This shows that $C\subset C(L)$, therefore $LC(L)=N(L)$. Clearly $L\cap C(L)=\{1\}$, and $C(L)$ is normal in $N(L)$. Hence $N(L)=L\times C(L)$ as was to be shown.

LEMMA 3.6. *The group G cannot contain a proper normal subgroup of odd index.*

Proof. Suppose G_0 is a proper normal subgroup of G of odd index. By assumption, the Sylow 2-groups of G_0 are not cyclic. If S_2 is a normal Sylow 2-group of G_0, then S_2 is characteristic in G_0, hence normal in G, contrary to hypothesis. Therefore by the induction hypothesis, $G_0=L\times D_0$, where L is isomorphic to $SL(2,2^a)$ and $|D_0|$ is odd. Since L is the subgroup of G_0 generated by all involutions in G_0, L is characteristic in G_0 and therefore normal in G. Lemma 3.5 now yields that $G=L\times D$ contrary to assumption.

LEMMA 3.7. *The group G cannot contain a proper normal subgroup of odd order.*

Proof. Let G_0 be a minimal proper normal subgroup of G with $|G_0|$ odd. By Lemma 3.6, G contains no proper normal subgroup of odd index.

Let S_2 be a Sylow 2-group of G, S_2G_0 is a group. If $G=S_2G_0$, then no two elements of S_2 are conjugate. By Lemma 3.1, this implies that S_2 contains only one involution. Since S_2 is abelian, it must be cyclic contrary to assumption, thus $S_2G_0\neq G$. It is clear that S_2G_0 cannot contain a subgroup L isomorphic to the simple group $SL(2,2^a)$. Since S_2 is not cyclic, the induction hypothesis implies that S_2 is a normal subgroup of S_2G_0, hence $S_2G_0=S_2\times G_0$. This shows that every involution is in $C(G_0)$. If G_0 is not in the center of G, then $C(G_0)$ is a proper normal subgroup of odd index, which is contrary to what has been assumed. Therefore G_0 is in the center of G.

If $x^2\in G_0$, x not in G_0, then $x=ux_0$, where $x_0\in G_0$ and u is an involution. If $yxy^{-1}G_0=xG_0$, then $yuy^{-1}G_0=uG_0$ since u is a power of x. Therefore $yuy^{-1}=u$ is the unique involution in uG_0. If $u\in S_2$, then $y\in C(S_2)$, hence the group G/G_0 satisfies condition (*). If the Sylow 2-group of G/G_0 is normal in G/G_0, then $S_2\times G_0$ is normal in G, thus S_2 is normal in G which

was assumed not to be the case. Since S_2 is not cyclic, the induction assumption implies that $G/G_0 = L \times D$, where L is isomorphic to $SL(2, 2^a)$ and $|D|$ is odd. The assumption that G contains no proper normal subgroup of odd index implies that $G/G_0 = L$.

Since G_0 is in the center of G, every subgroup of G_0 is normal in G. The minimality of G_0 now implies that G_0 has prime order p. Let S_p be a Sylow p-group of G. Suppose $S_p \neq G_0$, therefore p divides $(2^a + 1)$ or $(2^a - 1)$. In either case S_p/G_0 is a cyclic group, hence S_p is abelian. If $S_p = G_0$, then S_p is also abelian. Since G_0 is in the center of G, it is certainly in the center of $N(S_p)$. A theorem of Grün ([7] p. 173) now implies that G has a normal subgroup of odd index p, which is impossible. This completes the proof of the lemma.

The following two cases will now be treated separately. Case I will be taken care of in the next section by using a result from a previous paper [5]. Case II will be handled in Sections 5 and 6.

Case I. *For every y in $C(S_2) - \{1\}$, $C(y) \subset N(S_2)$.*

Case II. *There exists an element y in $C(S_2) - \{1\}$ such that $C(y)$ is not contained in $N(S_2)$.*

4. The proof for Case I. Throughout this section it will be assumed that G satisfies the hypothesis of Section 3. Furthermore, if $y \in C(S_2) - \{1\}$, then $C(y) \subset N(S_2)$. It will be shown that these assumptions lead to a contradiction.

Let $|C(S_2)| = c$, $|N(S_2)| = n = qc$.

LEMMA 4.1. $C(S_2) \cap xC(S_2)x^{-1} = \{1\}$ *unless* $x \in N(S_2)$.

Proof. If $y \in C(S_2) \cap xC(S_2)x^{-1}$, then both S_2 and xSx^{-1} are contained in $C(y)$. By assumption either $y = 1$ or $C(y) \subset N(S_2)$. Therefore if $y \neq 1$, S_2 and xS_2x^{-1} are both contained in $N(S_2)$, hence $xS_2x^{-1} = S_2$ and $x \in N(S_2)$.

LEMMA 4.2. *If y is real, $y^2 \neq 1$, then all elements in $C(y)$ are real.*

Proof. By Lemma 3.1, $|C(y)|$ is odd. Let u be an involution in $C^*(y)$, then $\{u\}$ is a Sylow 2-group of $C^*(y)$. Let $C_0 = C(u) \cap C^*(y)$. If $z \in C_0 \cap xC_0x^{-1}$ for some $x \in C^*(y)$, then $z \in C(S_2) \cap xC(S_2)x^{-1}$, where S_2 is the Sylow 2-group of G which contains u. If $z \neq 1$, then Lemma 4.1 implies that $x \in N(S_2)$. Therefore,

$$x \in N(S_2) \cap C^*(y) \subset C(u) \cap C^*(y) = C_0.$$

Consequently $C_0 \cap xC_0x^{-1} = \{1\}$ unless $x \in C_0$. This implies that no element in C_0 commutes with any element of $C^*(y) - C_0$, consequently $C_0 \cap C(y) = \{1\}$. Therefore by [1] Theorem 4B, every element in $C(y)$ is real.

LEMMA 4.3. *If* $y \in C(S_2) - \{1\}$, *then* $C(y) \subset C(S_2)$.

Proof. If $xy = yx$, $x \notin C(S_2)$, then by assumption $x \in N(S_2)$. Hence it may be asumed that $x^p \in C(S_2)$ for some prime p.

For any involution u not in S_2, $\{x\}C(S_2)u$ is a union of p cosets of $C(S_2)$ and therefore contains p involutions by Lemma 3.2. Let u, v be distinct involutions in $\{x\}C(S_2)u$. Then $x_0 = uv$ is a real element in $\{x\}C(S_2)$. By Lemmas 3.1 and 4.2, $C(x_0) \cap C(S_2) = \{1\}$. Therefore $\{x\}C(S_2) \cap C(x_0) = \{x_0\}$ and has order p. Since $\{x_0\}$ is isomorphic to its image in $\{x\}C(S_2)/C(S_2)$, no two elements in $\{x_0\}$ are conjugate in $\{x\}C(S_2)$. Therefore

$$N(\{x_0\}) \cap \{x\}C(S_2) = C(x_0) \cap \{x\}C(S_2) = \{x_0\}.$$

This implies that $\{x\}C(S_2)$ is a Frobenius group, consequently no element in $C(S_2) - \{1\}$ can commute with any element of $\{x\}C(S_2) - C(S_2)$. This contradicts the choice of x and proves the lemma.

LEMMA 4.4. *The group G cannot satisfy all the assumptions stated at the beginning of this section.*

Proof. No element of $C(S_2) - \{1\}$ commutes with any element of $N(S_2) - C(S_2)$, hence $(q, c) = 1$. Lemma 4.1 now implies that $g = qc(1 + kc)$ for some integer $k \geqq 1$. It follows from Lemma 3.4 that

$$(g/qc - 1)(q - 1) \leqq qc - c = (q - 1)c.$$

Therefore $g \leqq qc(1 + c)$. Since $k \neq 0$, $g = qc(1 + c)$. It is now easy to see that G is a doubly transitive permutation group on the $(1 + c)$ conjugates of $C(S_2)$. The next step is to show that no non-trivial permutation leaves three conjugates of $C(S_2)$ fixed. In other words if N_1, N_2, N_3 are distinct subgroups conjugate to $N(S_2)$, then $N_1 \cap N_2 \cap N_3 = \{1\}$.

As $(q, c) = 1$, Lemma 4.1 implies that the order of $N_1 \cap N_2$ divides q, and therefore no element of $N_1 \cap N_2$ commutes with any involution G. Since N_1 is its own normalizer, and every coset of N_1 contains involutions, there exist involutions u, v such that $N_2 = uN_1u$, $N_3 = vN_1v$. Therefore $uN_2u = N_1$, and $vN_3v = N_1$, hence u is in the normalizer of $N_1 \cap N_2$, and v is in the normalizer of $N_1 \cap N_3$. Consequently $\{u\}(N_1 \cap N_2)$ is a group. If y is an element of $N_1 \cap N_2$, then $u \in C^*(y)$, since no element of $N_1 \cap N_2$ commutes

with any involution. Therefore if $y \in N_1 \cap N_2 \cap N_3$, both u and v are contained in $C^*(y)$, hence $uv \in C(y) \subset N_1$ unless $y = 1$. Consequently $uN_1u = vN_1v$ if $y \neq 1$. This is impossible, therefore $N_1 \cap N_2 \cap N_3 = \{1\}$.

The group G now satisfies the hypotheses of [5] Theorem 1. Hence c is a power of some prime. Since c is even, this implies that $C(S_2) = S_2$. The group S_2 is abelian of order 2^a, therefore by [5] Theorem 1, $q \geqq \frac{1}{2}(2^a - 1)$. Since q divides $2^a - 1$, this implies that $q = 2^a - 1$. A theorem of H. Zassenhaus (see [8]) can now be applied to show that G is isomorphic to $SL(2, 2^a)$. This is contrary to the assumption stated at the beginning of this section.

5. Case II. The group Q. Throughout this section it will be assumed that the group G satisfies the hypothesis stated at the beginning of Section 3. Furthermore, there exists an element y in $C(S_2) - \{1\}$ with the property that $C(y)$ is not contained in $N(S_2)$. Consequently the results of Section 3 may be used.

The group $C(y)$ contains S_2, but S_2 is not normal in $C(y)$. Since y has odd order, Lemma 3.7 implies that $C(y) \neq G$. Consequently by hypothesis $C(y) = L \times D_0$, where $|D_0|$ is odd, and L is isomorphic to $SL(2, 2^a)$. By Lemma 3.5, $N(L) = L \times D$ for some group D of odd order. Let $|D| = d$, then $d > 1$.

By Lemma 2.1, S_2 is the direct product of groups of order two, and $N(S_2)$ contains a cyclic subgroup Q of order $q = (2^a - 1)$ in which every element is real, $Q \cap C(S_2) = \{1\}$. By Lemma 3.1,

$$[N(S_2) : C(S_2)] = q = (2^a - 1).$$

LEMMA 5.1. $N(L) = N(D) = N \times D$.

Proof. As G contains no normal subgroup of odd order, $N(D) \neq G$. Hence by induction $N(D) = L \times D_0$ for some group D_0 of odd order, $L \times D \subset L \times D_0$. Since L is normal in $L \times D_0$, $L \times D_0 \subset N(L) = L \times D$.

LEMMA 5.2. *If x is not contained in $N(D)$, then $xDx^{-1} \cap D = \{1\}$. If D_0 is a subgroup of D of order greater than one, then $N(D_0) \subset N(D)$.*

Proof. Let $D_0 = D \cap xDx^{-1}$, suppose $D_0 \neq 1$. Therefore L, xLx^{-1} are both contained in $N(D_0)$. As G has no proper normal subgroup of odd order $N(D_0) \neq G$, hence by induction $N(D_0)$ contains exactly $(2^a + 1)$ Sylow 2-groups of G. These must lie in $L \cap xLx^{-1}$, hence $L = xLx^{-1}$ and $x \in N(L) = N(D)$. The second statement is clear.

LEMMA 5.3. $C(S_2) = S \times H$. *Either $H = D$ or H is a Frobenius group whose regular subgroup M has order m and $|H| = md$, $(m, d) = 1$.*

Proof. By Lemma 3.1, $C(S_2)=S_2\times H$. Lemma 5.1 implies that

$$N(D)\cap C(S_2)=(L\times D)\cap C(S_2)=S_2\times D,$$

therefore $N(D)\cap H=D$. If $D\neq H$, then by Lemma 5.2, D intersects no conjugate non-trivially and is its own normalizer in H. Therefore H has the required properties (see [3] p. 334).

LEMMA 5.4. *If $M\neq\{1\}$, then $N(M)=N(S_2)$. If M_0 is any non-empty subset of M, $M_0\neq\{1\}$, then $N(M_0)\subset N(S_2)$.*

Proof. Since $(m,2^ad)=1$ and M is normal in $C(S_2)$, it follows that M is characteristic in $C(S_2)$, hence normal in $N(S_2)$, $N(S_2)\subset N(M)$. Let M_0 be a non-empty subset of M other than $\{1\}$, suppose $N(M_0)$ is not contained in $N(S_2)$. Since $S_2\subset N(M_0)$, the induction assumption implies that there is a subgroup L_0 of G, isomorphic to $SL(2,2^a)$ such that $N(M_0)=L_0\times M_1$, where $M_0\subset M_1$. Since $M_0\neq H$, it follows from Lemmas 5.1 and 5.3 applied to $N(L_0)$ that the normal subgroup of H generated by M_0 is equal to H. This is not the case since $M_0\subset M$. Consequently, $N(M_0)\subset N(S_2)$ as was to be shown.

LEMMA 5.5. *If $y\in Q-\{1\}$, then $C(y)=Q\times D$, and $C^*(y)=C^*(Q)=\{u\}(Q\times D)$ for some involution u in $C(D)$.*

Proof. By Lemma 3.3, $C(y)=QDM_0$, where $M_0\subset M$ and no element of DM_0 is real. Consequently no coset of DM_0 in $C^*(y)$ can contain more than one involution. There is no involution in $C(y)$, hence $C^*(y)$ contains at most q involutions. Lemma 2.1 now implies that $C^*(y)$ contains exactly q involutions, all of these lie in $C(D)$. If $M_0\neq\{1\}$, then $xDx^{-1}\neq D$ for $x\in M_0-\{1\}$ and by the same argument as above every involution lies in $C(xDx^{-1})$. Let u be an involution in $C^*(y)$, then $\{D,xDx^{-1}\}\subset C(u)$. The group generated by D and xDx^{-1} contains some element $y\in M_0-\{1\}$, therefore $u\in C(y)$. This contradicts Lemma 5.4 since u is not in $N(S_2)$.

LEMMA 5.6. *If x is not in $N(Q)$, then $Q\cap xQx^{-1}=\{1\}$.*

Proof. Let $Q_0=Q\cap xQx^{-1}$, suppose $Q_0\neq\{1\}$. Since Q is cyclic, so is Q_0, let $Q_0=\{y\}$. By Lemma 5.5

$$C^*(y)=\{u\}(Q\times D).$$

Hence all the involutions in $C^*(y)$ lie in $\{u\}Q$, by a similar argument they also all lie in $\{u\}xQx^{-1}$. Therefore all the elements which are a product of two involutions in $C^*(y)$ lie in $Q\cap xQx^{-1}=Q_0$. Since $C^*(y)$ contains q

involutions, it contains at least q distinct elements which are the product of two involutions, hence $|Q_0| \geqq q$. Therefore $Q_0 = Q = xQx^{-1}$ and $x \in N(Q)$.

LEMMA 5.7. *No element of $Q - \{1\}$ commutes with any element of $M - \{1\}$. $(q, m) = 1$.*

Proof. QM has a normal subgroup M, $Q \cap M = \{1\}$, and Q intersects no conjugate non-trivially. By Lemma 5.5

$$N(Q) \cap QM \subset N(Q) \cap N(S_2) = C(Q) = Q \times D.$$

Hence

$$N(Q) \cap QM \subset (Q \times D) \cap QM = Q.$$

Therefore QM is a Frobenius group with regular subgroup M. This proves the lemma.

LEMMA 5.8. *Q is contained in exactly two conjugates N_1, N_2 of $N(S_2)$, $N_1 \cap N_2 = Q \times D = C(Q)$. $N(Q \times D) = C^*(Q)$.*

Proof. Let $N_1 = N, N_2, \cdots, N_k$ be all the conjugates of N containing Q. Since N is its own normalizer it follows from Lemma 3.2 that for each $i = 2, \cdots, k$, there exist q involutions $u_{i1}, \cdots, u_{iq}$ such that $u_{ij}N_1u_{ij} = N_i$. Therefore $u_{ij}N_iu_{ij} = N_1$, hence

$$u_{ij}(N_1 \cap N_i)u_{ij} = N_1 \cap N_i.$$

Lemma 3.3 implies that

$$Q \times D = C(Q) \subset N_1 \cap N_i.$$

Since $(qd, 2^am) = 1$,

$$N_1 \cap N_i \cap (M \times S_2) = (M \times S_2) \cap u_{ij}(M \times S_2)u_{ij} = M_0.$$

Since u_{ij} is in its normalizer, M_0 must have odd order, hence $M_0 \subset M$. Therefore by Lemma 5.4, $M_0 = \{1\}$, consequently

$$Q \times D = N_1 \cap N_i, \qquad i = 2, \cdots, k.$$

Therefore each of the involutions u_{ij} is in $N(Q \times D)$, $N(Q \times D)$ acts as a transitive permutation group on $N_1, \cdots, N_k$. The subgroup leaving N_1 fixed is $N(Q \times D) \cap N_1$. Since no element whose order divides 2^a can be in this subgroup, $N(Q \times D) \cap N_1 = Q \times D$. Hence $|N(Q \times D)| = kqd$, and $N(Q \times D)$ contains at least $q(k-1)$ involutions. The Sylow 2-group of $N(Q \times D)$ is obviously not normal, since elements of Q are conjugate to their inverse in $N(Q \times D)$. Hence the induction assumption implies that

there is only one class of involutions in $N(Q \times D)$. Let u be an involution in $C^*(Q)$, then $\{u\} \times D \subset C(u)$, hence $|C(u)| \geqq 2d$, therefore $N(Q \times D)$ contains at most $kqd/2d$ involutions. Therefore $q(k-1) \leqq \frac{1}{2}qk$, hence $k \leqq 2$. Since $N(Q \times D) \neq Q \times D$, this implies that $k = 2$ which suffices to prove all the statements of the lemma.

LEMMA 5.9. *If $y \in Q$, $xyx^{-1} \in Q \times D$, then $xyx^{-1} = y$ or y^{-1}. In particular, $N(Q) = C^*(Q)$.*

Proof. If $x \in N(S_2)$, then $xyx^{-1}S_2M = yS_2M$, since S_2M is a normal subgroup of $N(S_2)$ and yS_2M is in the center of $N(S_2)/S_2M$. Therefore $xyx^{-1} = y$, as $(Q \times D) \cap S_2M = \{1\}$.

If x is not $N(S_2)$, then $x = x_0u$, where $x_0 \in C(S_2)$, and u is an involution not in S_2. Therefore

$$y \in N(S_2) \cap x^{-1}N(S_2)x = N(S_2) \cap uN(S_2)u.$$

By Lemmas 3.3 and 5.8, this implies that

$$Q \times D \subset C(y) \subset N(S_2) \cap uN(S_2)u = Q \times D,$$

and therefore $u \in N(Q \times D) = C^*(Q)$. Consequently uyu^{-1}, $uyu = y^{-1}$ and $xyx^{-1} = x_0y^{-1}x_0^{-1}$. Since $x_0 \in N(S_2)$, the first part of the lemma implies that $x_0y^{-1}x_0^{-1} = y^{-1}$.

LEMMA 5.10. *If $y \in Q \times D$ and y is real, then $y \in Q$.*

Proof. If $x^{-1}yx = y^{-1}$, then x is not in $N(S_2)$, hence $x = x_0u$, where $x_0 \in C(S_2)$ and u is an involution not in S_2. Therefore

$$y \in N(S_2) \cap x^{-1}N(S_2)x = N(S_2) \cap uN(S_2)u.$$

By Lemmas 3.3 and 5.8, this implies that

$$C(y) \subset N(S_2) \cap uN(S_2)u = Q \times D$$

and therefore $u \in N(Q \times D) = C^*(Q)$. Consequently

$$x_0^{-1}yx_0 = ux^{-1}yxu = uy^{-1}u \in Q \times D.$$

Since S_2M is a normal subgroup of $N(S_2)$ and $(2^a m, qd) = 1$, this implies that there is an element $x_1 \in Q \times D$ such that $x_1^{-1}yx_1 = x_0^{-1}yx_0$. Therefore $x_2 = x_1u \in C^*(Q)$ and

$$x_2^{-1}yx_2 = ux_1^{-1}yx_1u = ux_0^{-1}yx_0u = x^{-1}yx = y^{-1}.$$

However Lemma 5.5 implies that the only elements in $Q \times D$ which are conjugate to their inverses in $C^*(Q)$ are in Q. Hence $y \in Q$.

LEMMA 5.11. *If* $y \in Q \times D$ *and* $C(y)$ *is not contained in* $N(S_2)$, *then* $y \in D$.

Proof. By Lemma 3.3 there is an involution $v \in C(y)$, hence $y \in C(v)$. Since the order of y is relatively prime to $2^a m$ by Lemmas 5.3 and 5.7, y is contained in a subgroup D_0 of $C(v)$ of order d. All subgroups of order d are conjugate to each other in $C(v)$, hence D_0 is conjugate to D. Since $Q \subset C(y)$, Lemma 5.2 implies that $Q \subset N(D_0)$. As D_0 contains no real elements and every element of Q is real, $Q \cap D_0 = \{1\}$. Since $N(D_0) = L_0 \times D_0$, where L_0 is conjugate to L, there exists a subgroup Q_0 of L_0 of order q such that $QD_0 = Q_0 \times D_0$. Lemma 5.9 implies that $Q = Q_0$, hence $D_0 \subset C(Q)$, therefore

$$Q \times D = C(Q) = Q \times D_0.$$

There are q involutions in $C^*(Q) \cap C(D)$ and q in $C^*(Q) \cap C(D_0)$. By Lemma 5.5, there are only q involutions in $C^*(Q)$. Therefore for any involution u in $C^*(Q)$, $\{D, D_0\} \subset C(u)$. Since $|C(u) \cap C^*(Q)| = 2d$, this implies that

$$\{u\} \times D = C(u) \cap C^*(Q) = \{u\} \times D_0.$$

Consequently $D = D_0$ is the set of all elements of odd order in $C(u) \cap C^*(Q)$. Therefore $y \in D_0 = D$.

LEMMA 5.12. $(q, d) = 1$ *or* $m = 1$. *If* $m \neq 1$, *every Sylow subgroup of* D *is cyclic.*

Proof. We will first show that if $m \neq 1$, then no element of $(Q \times D) - \{1\}$ commutes with any element of $M - \{1\}$. Suppose this is not the case and there is an element $y \in (Q \times D) - \{1\}$ such that y commutes with some element of $M - \{1\}$, $C(y) \cap M \neq \{1\}$. By Lemma 5.2, $y \in (Q \times D) - D$, hence Lemma 5.11 implies that $C(y) \subset N(S_2)$. Let u be an involution in $C^*(Q)$, then

$$y \in Q \times D \subset uN(S_2)u \neq N(S_2),$$

and $uDu = D$. Since

$$M \cap uNu = M \cap uMu = \{1\}$$

by Lemma 5.4, $C(y)$ is not contained in uNu. Hence by Lemma 5.11, $y \in D$ which contradicts an earlier statement.

It now follows from [4] Lemma 2.1 that QDM is a Frobenius group whose regular subgroup is M. A theorem of Burnside ([3] p. 336) implies that every Sylow group of $Q \times D$ is cyclic, since qd is odd. Therefore $(q, d) = 1$.

LEMMA 5.13. $g = q2^a dm(1 + 2^a m)$.

Proof. It follows from Lemma 5.10 that the only real elements in $Q \times D$ are in Q. There are $2^a m$ subgroups of $N(S_2)$ conjugate to $Q \times D$. For each subgroup D_0 conjugate to D, there are 2^a conjugates of Q in $C(D_0)$. If $m \neq 1$, and $x \in M - \{1\}$, then $xQDx^{-1} \cap QD = \{1\}$, since otherwise by Lemma 5.12, the intersection would have to contain elements of either D or Q which is not the case. A simple computation now yields that $N(S_2)$ contains at least $2^a dm + m2^a(qd - q - d + 1)$ non-real elements. Hence $N(S_2)$ contains at most $2^a m(q-1)$ real elements.

Since the intersection of $S_2 M$ with any conjugate distinct from $S_2 M$ has order one, it is easy to show that $g = q2^a md(1 + k2^a m)$, where k is an integer with $k \geqq 1$. It follows from Lemma 3.4 that

$$(g/q2^a md - 1)(q-1) \leqq 2^a m(q-1).$$

Hence $g \leqq q2^a md(1 + 2^a m)$, therefore $g = q2^a md(1 + 2^a m)$.

We will now state an omnibus lemma which summarizes most of the results in this section that are needed for the proof of Theorem 1.

LEMMA 5.14.

(i) $m \neq 1$, $(q, d) = 1$.

(ii) *Every Sylow subgroup of* D *is cyclic.*

(iii) *If* $y \in (Q \times D) - D$, *then* $xyx^{-1} \in Q \times D$ *only if* $x \in C^*(Q)$.

(iv) $N(Q \times D) = C^*(Q) = \{u\}(Q \times D)$ *for some involution* u.

(v) *If* $y \in (S_2 \times D) - D - S_2$, *then* $xyx^{-1} \in S_2 \times D$ *only if* $x \in N(S_2 \times D)$.

(vi) $N(S_2 \times D) = QS_2 \times D$.

Proof. If $m = 1$, then $g = |L \times D|$, hence $G = L \times D$ contrary to assumption. Lemma 5.12 implies that $(q, d) = 1$ and that every Sylow subgroup of D is cyclic.

If $y \in (Q \times D) - D$, then since $(q, d) = 1$, some power y_0 of y is in $Q - \{1\}$. Hence $xy_0x^{-1} \in Q \times D$, therefore by Lemma 5.9, $x \in C^*(Q)$. Statement (iv) follows from Lemmas 5.5 and 5.8. Statements (v) and (vi) are immediate consequences of Lemma 2.1.

6. Completion of the proof for Case II. Throughout this section the group G will be assumed to satisfy the hypothesis stated at the beginning of

Section 3. Furthermore, there exists an element y in $C(S_2) - \{1\}$ with the property that $C(y)$ is not contained in $N(S_2)$. Consequently the results of Sections 3 and 5 can be used.

Our aim is to reach a contradiction. Once this has been achieved it will, together with Lemma 4.4, complete the proof of Theorem 1 in all cases.

Let D' denote the commutator subgroup of D. By Lemma 5.14 and the results of [7] page 175, $(|D'|, [D: D']) = 1$, and D' and D/D' are cyclic groups, $D \neq D'$. Let p be a prime which divides $[D: D']$, let D_p be a Sylow p-group of G. It is a simple consequence of Lemma 5.2 that no two elements in D_p are conjugate, and every subgroup of D_p is contained in the center of its normalizer. By Lemma 2.1, L contains a cyclic subgroup K of order $(2^a + 1)$. Let K_p denote the Sylow p-group of K, let $K_p{}^0$, $D_p{}^0$ denote the subgroup generated by all elements of order p in K_p, D_p respectively. The major portion of this section is devoted to showing that $N(K_p{}^0 \times D_p{}^0) \subset L \times D$.

Let D_1 be the normalizer of $D_p{}^0$ in D, $|D_1| = d_1$. Let S_p be a Sylow p-group of G which contains $K_p \times D_p$.

LEMMA 6.1. $K_p{}^0 \neq \{1\}$.

Proof. If $K_p{}^0 = \{1\}$, then by Lemma 5.2, $N(D_p) \subset L \times D$. Hence D_p is a Sylow p-group of G. Since D_p is contained in the center of its normalizer, a theorem of Burnside ([3] page 327) implies that G contains a normal subgroup of index p, this contradicts Lemma 3.6.

LEMMA 6.2.

$$N(K_p{}^0 \times D_p{}^0) \cap (L \times D) = \{u\} (K \times D_1) = N(K_p{}^0 \times D_p{}^0) \cap C(D_p{}^0),$$

where u is an involution in $C(D_1)$. Furthermore, if

$$[N(K_p{}^0 \times D_p{}^0) : N(K_p{}^0 \times D_p{}^0) \cap (L \times D)] = h,$$

then $h \leqq p$. $h = p$ if and only if $K_p \times D_p \neq S_p$, in that case there are exactly p subgroups of $K_p{}^0 \times D_p{}^0$ which are conjugate to $D_p{}^0$, and they are all conjugate to $D_p{}^0$ in $N(K_p{}^0 \times D_p{}^0)$.

Proof. The first statement follows immediately from Lemma 2.1. Since $K_p{}^0$, $D_p{}^0$ are both cyclic, $K_p{}^0 \times D_p{}^0$ contains exactly $p + 1$ subgroups of order p. Every element of $K_p{}^0$ is real, no element of $D_p{}^0 - \{1\}$ is real, therefore there are at most p subgroups conjugate to $D_p{}^0$ in $K_p{}^0 \times D_p{}^0$. If $xD_p{}^0x^{-1} = yD_p{}^0y^{-1}$, then $x^{-1}y \in N(D_p{}^0)$, therefore $x^{-1}y \in L \times D$. This shows that $h \leqq p$.

If $h = p$, then $K_p \times D_p$ is not a Sylow p-group of $N(K_p{}^0 \times D_p{}^0)$, there-

fore $K_p \times D_p \neq S_p$. Conversely suppose that $K_p \times D_p \neq S_p$, then $K_p \times D_p$ is a proper subgroup of a Sylow p-group of $N(K_p \times D_p)$. Since $K_p{}^0 \times D_p{}^0$ is a characteristic subgroup of $K_p \times D_p$, this implies that $K_p \times D_p$ is a proper subgroup of a Sylow p-group of $N(K_p{}^0 \times D_p{}^0)$. Therefore h is divisible by p, hence $h = p$. As $N(D_p{}^0) \subset L \times D$, $D_p{}^0$ has p conjugates in $K_p{}^0 \times D_p{}^0$.

LEMMA 6.3. *Suppose that for some element x in $D_p{}^0 - \{1\}$ there exists an element z with $zxz^{-1} \in K_p{}^0 \times D_p{}^0$. Then there exists an element y such that $zxz^{-1} = yxy^{-1}$, $z^{-1}y \in L \times D_1$ and $y \in N(K_p{}^0 \times D_p{}^0)$.*

Proof. Assume first that $K_p \times D_p = S_p$. Then ([7] page 169) there exists an element $y \in N(K_p \times D_p)$ such that $yxy^{-1} = zxz^{-1}$. Therefore $z^{-1}y \in C(x)$, hence $z^{-1}y \in L \times D_1$. As $K_p{}^0 \times D_p{}^0$ is a characteristic subgroup of $K_p \times D_p$, $y \in N(K_p{}^0 \times D_p{}^0)$.

Suppose now that $K_p \times D_p \neq S_p$. By Lemma 6.2 there exists an element $y \in N(K_p{}^0 \times D_p{}^0)$ such that $zD_p{}^0z^{-1} = yD_p{}^0y^{-1}$, hence $z^{-1}y \in N(D_p{}^0)$. Therefore $z^{-1}y \in C(D_p{}^0)$, since $D_p{}^0$ is in the center of $N(D_p{}^0)$. Consequently $z^{-1}y \in L \times D_1$ and $zxz^{-1} = yxy^{-1}$.

Let λ_0 be the trivial character of $L \times D_p{}^0$, let $\lambda \neq \lambda_0$ be a character of $L \times D_p{}^0$ whose representation contains L in its kernel, hence $\lambda(1) = 1$. Let $\lambda_0{}^*, \lambda^*$ be the characters of G induced by λ_0, λ respectively, and let $\alpha = \lambda_0{}^* - \lambda^*$. Let β be the character of G induced by the trivial character of $N(S_2)$.

We wish to show that $h = 1$, where h has the same meaning as in Lemma 6.2. This wil be done by using arithmetical properties of α and β which will be established in the next few lemmas.

LEMMA 6.4. *$\alpha(x)\beta(x) = 0$ unless x is conjugate to an element in one of the following three sets: $(Q \times D_p{}^0) - Q - D_p{}^0$, $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$, $D_p{}^0 - \{1\}$.*

Proof. It follows immediately from the definition of α, that $\alpha\beta$ vanishes on elements not conjugate to an element of $(L \times D_p{}^0) - L$. Hence by Lemma 2.1, if $\alpha(x)\beta(x) \neq 0$, then $x^{pq} = 1$, or $x^{p2^a} = 1$, or $x^{p(2^a+1)} = 1$.

If $x^{pq} = 1$, then x is conjugate to an element of $Q \times D_p{}^0$. Since $\alpha\beta$ vanishes on elements whose order divides q, x is conjugate to an element of $(Q \times D_p{}^0) - Q - D_p{}^0$, or of $D_p{}^0 - \{1\}$. A similar argument shows that if $x^{p2^a} = 1$, then x is conjugate to an element of $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$ or of $D_p{}^0 - \{1\}$.

Suppose $x^{p(2^a+1)} = 1$, then x is conjugate to an element of $K \times D_p{}^0$. Since $\beta(x) \neq 0$, x is conjugate to an element of D. Hence by taking conjugates it may be assumed that there exists an element y with

$$x \in D \cap y(K \times D_p{}^0)y^{-1}.$$

Hence $x^p \in yKy^{-1}$, therefore x^p is real. Since $x^p \in D$, this implies that $x^p = 1$. Therefore x is conjugate to an element D_p. Since D_p is cyclic, this finally implies that x is conjugate to an element of $D_p{}^0$. As $\alpha(1) = 0$, $x \neq 1$.

LEMMA 6.5.

$$\alpha(x)\beta(x) = \begin{cases} (2d_1/p)\{1-\lambda(x)\} & \text{if } x \in (Q \times D_p{}^0) - Q - D_p{}^0 \\ (d_1/p)\{1-\lambda(x)\} & \text{if } x \in (S_2 \times D_p{}^0) - S_2 - D_p{}^0 \\ (2^a+1)\alpha(x) & \text{if } x \in D_p{}^0 - \{1\}. \end{cases}$$

Proof. It follows from Lemma 5.14 that if x and yxy^{-1} are both in $(Q \times D_p{}^0) - Q - D_p{}^0$, or in $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$, then $y \in N(Q \times D_p{}^0)$ or $y \in N(S_2 \times D_p{}^0)$ respectively. Lemma 5.14 also implies that $N(Q \times D_p{}^0) = \{u\}(Q \times D_1)$, where u is an involution in $C(D) \cap C^*(Q)$, and $N(S_2 \times D_p{}^0) = QS_2D_1$. Since $D_p{}^0$ is in the center of both these groups, the values of $\alpha(x)\beta(x)$ for x in $(Q \times D_p{}^0) - Q - D_p{}^0$ or x in $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$ are easily computed to be those in the statement of the Lemma. If $x \in D_p{}^0 - \{1\}$, then $C(x) = L \times D_1$, therefore it follows from the definition of β, that $\beta(x) = (2^a+1)$.

LEMMA 6.6. *Consider the class of subsets of G which consists of all sets conjugate to one of the sets $(Q \times D_p{}^0) - Q - D_p{}^0$, $(S_2 \times D_p{}^0) - Q - D_p{}^0$, $D_p{}^0 - \{1\}$. No two of the sets in this class have an element in common.*

Proof. If x is conjugate to an element of $(Q \times D_p{}^0) - Q - D_p{}^0$, then the order of x is odd and has a factor in common with q. If x is conjugate to an element of $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$, then the order of x is even. If x is conjugate to an element of $D_p{}^0 - \{1\}$, then the order of x is relatively prime to $2q$. Hence a set in this class can intersect only its conjugates. It follows from Lemma 5.14 that neither $(Q \times D_p{}^0) - Q - D_p{}^0$ nor $(S_2 \times D_p{}^0) - S_2 - D_p{}^0$ has any element in common with any of its conjugates. Since $|D_p{}^0| = p$, the statement is trivially true for $D_p{}^0 - \{1\}$.

LEMMA 6.7. *$(Q \times D_p{}^0) - Q - D_p{}^0$ has $g/2qd_1$ conjugates, $(S_2 \times D_p{}^0) - S - D_p{}^0$ has $g/q2^ad_1$ conjugates, $D_p{}^0 - \{1\}$ has $g/q2^a(2^a+1)d_1$ conjugates.*

Proof. By Lemma 5.14, $|N(Q \times D_p{}^0)| = 2qd_1$, $|N(S_2 \times D_p{}^0)| = q2^ad_1$. By Lemma 5.2, $N(D_p{}^0) = L \times D_1$, hence $|N(D_p{}^0)| = q2^a(2^a+1)d_1$.

LEMMA 6.8. *The expression*

$$(1) \qquad (1/q2^a)[(1/d_1)\sum_{D_p{}^0}\alpha(x) - (1/p)\sum_{D_p{}^0}\{1-\lambda(x)\}]$$

is a rational integer.

Proof. It follows from the definition of α and β, that $(\alpha,\bar{\beta})_G$ is a rational integer. Since $\alpha(x)=0$ for $x\in S_2$ or $x\in Q$, Lemmas 6.4, 6.6 and 6.7 imply that

$$(2)\qquad \begin{aligned}(\alpha,\bar{\beta})_G = (1/g)[(g/2qd_1)&\sum_{(Q\times D_p^0)-D_p^0}\alpha(x)\beta(x)\\ &+(g/q2^a d_1)\sum_{(S_2\times D_p^0)-D_p^0}\alpha(x)\beta(x)\\ &+(g/q2^a(2^a+1)d_1)\sum_{D_p^0}\alpha(x)\beta(x)].\end{aligned}$$

Lemma 6.5 can be used to compute each of the terms appearing in equation (2). We get

$$(3)\qquad \begin{aligned}&(1/2qd_1)\sum_{(Q\times D_p^0)-D_p^0}\alpha(x)\beta(x)\\ &\quad=(1/pq)\sum_{Q\times D_p^0}\{1-\lambda(x)\}-(1/pq)\sum_{D_p^0}\{1-\lambda(x)\},\end{aligned}$$

$$(4)\qquad \begin{aligned}&(1/q2^a d_1)\sum_{(S_2\times D_p^0)-D_p^0}\alpha(x)\beta(x)\\ &\quad=(1/q2^a p)\sum_{S_2\times D_p^0}\{1-\lambda(x)\}-(1/q2^a p)\sum_{D_p^0}\{1-\lambda(x)\},\end{aligned}$$

$$(5)\qquad (1/q2^a(2^a+1)d_1)\sum_{D_p^0}\alpha(x)\beta(x)=(1/q2^a d_1)\sum_{D_p^0}\alpha(x).$$

Since λ is a character of $S_2\times D_p^0$ whose representation contains S_2 in its kernel, it follows that

$$(6)\qquad (1/q2^a p)\sum_{S_2\times D_p^0}\{1-\lambda(x)\}=(1/qp)\sum_{D_p^0}\{1-\lambda(x)\}.$$

If equations (3), (4), (5), and (6) are now substituted in (2) we get that

$$(7)\qquad \begin{aligned}(\alpha,\beta)_G=(1/pq)&\sum_{Q\times D_p^0}\{1-\lambda(x)\}\\ &-(1/q2^a p)\sum_{D_p^0}\{1-\lambda(x)\}+(1/q2^a d_1)\sum_{D_p^0}\alpha(x).\end{aligned}$$

The first term on the right hand side of equation (7) is the inner product of the generalized character $1-\lambda$ of $Q\times D_p^0$ with the trivial character of $Q\times D_p^0$, hence it is a rational integer. Therefore the sum of the last two terms on the right hand side of equation (7) is a rational integer. This sum is precisely the expression in equation (1). This proves the lemma.

LEMMA 6.9. *For $x\in D_p^0$*

$$(8)\qquad \alpha(x)=(d_1/p)\sum_{i=1}^{h}\{1-\lambda(y_i x y_i^{-1})\}=(d_1/p)[h-\sum_{i=1}^{h}\lambda(y_i x y_i^{-1})],$$

where $y_1,\cdots,y_h$ are coset representatives of the h coset of $C(D_p^0)$ in $N(K_p^0\times D_p^0)$.

Proof. If $x = 1$, then both sides of equation (8) are zero. Assume $x \neq 1$, let $z_1, z_2, \cdots$ be a set of representatives of all the left cosets of $L \times D_1$ in G with $z_i x z_i^{-1} \in L \times D_p{}^0$. For each z_i, there exists an element $x_i \in L \times D_p{}^0$ such that $(z_i x_i) x (z_i x_i)^{-1} \in K_p \times D_p{}^0$, since $K_p \times D_p{}^0$ is a Sylow p-group of $L \times D_p{}^0$. As $K_p{}^0 \times D_p{}^0$ contains all elements of order p in $K_p \times D_p{}^0$, $(z_i x_i) x (z_i x_i)^{-1}$ is in $K_p{}^0 \times D_p{}^0$. Since $z_i x_i$ is in the same left coset of $L \times D_1$ as z_i, it may be assumed by a change of notation that $z_i x z_i^{-1} \in K_p{}^0 \times D_p{}^0$ for all z_i. Lemma 6.3 implies that for each z_i there exists an element y_i in the same left coset of $L \times D_1$ as z_i, such that $y_i x y_i^{-1} = z_i x z_i^{-1}$ and y_i is in $N(K_p{}^0 \times D_p{}^0)$. Let $w_1, w_2, \cdots$ be a system of coset representatives of $D_p{}^0$ in D_1. Then by Lemma 6.2 the set of elements $y_i w_j$ is a system of coset representatives of $\{u\}(K \times D_p{}^0)$ in $N(K_p{}^0 \times D_p{}^0)$, and

$$\alpha(x) = \sum_{i,j} \{1 - \lambda(y_i w_j x w_j^{-1} y_i^{-1})\} = (d_1/p) \sum_{i=1}^{h} \{1 - \lambda(y_i x y_i^{-1})\}. \tag{9}$$

LEMMA 6.10. $N(K_p{}^0 \times D_p{}^0) \subset L \times D$.

Proof. It follows from Lemma 6.2 that it is sufficient to show that $h = 1$ in Lemma 6.9. It is easily seen that

$$\sum_{D_p{}^0} \lambda(y_i x y_i^{-1}) = \sum_{y_i^{-1} D_p{}^0 y_i} \lambda(x). \tag{10}$$

The definition of λ implies that the restriction of λ to $K_p{}^0 \times D_p{}^0$ is a character of degree one whose kernel is $K_p{}^0$. Hence the restriction of λ to any proper subgroup of $K_p{}^0 \times D_p{}^0$, distinct from $K_p{}^0$, is not the trivial character of that subgroup. As every element of $K_p{}^0$ is real and no element of $D_p{}^0 - \{1\}$ is real, $K_p{}^0 \neq y_i^{-1} D_p{}^0 y_i$ for all y_i. Therefore equation (10) implies that

$$\sum_{D_p{}^0} \lambda(y_i x y_i^{-1}) = \sum_{y_i^{-1} D_p{}^0 y_i} \lambda(x) = 0. \tag{11}$$

If the value for α in equation (8) is substituted into equation (1), and equation (11) is taken into account,, we get by Lemma 6.8 that the expression

$$\begin{aligned} &(1/q2^a)[(h-1) + (1/p) \sum_{D_p{}^0} \lambda(x) - \sum_{i=1}^{h} \{(1/p) \sum_{D_p{}^0} \lambda(y_i x y_i^{-1})\}] \\ &= (h-1)/q2^a \end{aligned} \tag{12}$$

is a rational integer. However Lemmas 6.1 and 6.2 imply that $h \leqq p \leqq (2^a + 1)$, therefore $(h-1) \leqq 2^a$. This together with equation (12) finally shows that $h = 1$ which suffices to prove the lemma.

LEMMA 6.11. *The group G cannot satisfy all the assumptions stated at the beginning of this section.*

Proof. Since $K_p{}^0 \times D_p{}^0$ is a characteristic subgroup of $K_p \times D_p$, it follows from Lemma 6.10, that $N(K_p \times D_p) \subset L \times D$. Therefore $S_p = K_p \times D_p$ is a Sylow p-group of $N(S_p)$, hence S_p is an abelian Sylow p-group of G. By Lemma 6.1, $D_p{}^0$ is in the center of $N(S_p)$, therefore ([7] page 173) G contains a normal subgroup of index p. This contradicts Lemma 3.6 and completes the proof.

As was shown earlier, Lemma 6.11 is sufficient to complete the proof of Theorem 1.

CORNELL UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] R. Brauer and K. A. Fowler, "On groups of even order," *Annals of Mathematics*, vol. 62 (1955), pp. 565-583.

[2] ———, M. Suzuki and G. E. Wall, "A characterization of the one dimensional unimodular projective groups over finite fields," *Illinois Journal of Mathematics*, vol. 2 (1958), pp. 718-745.

[3] W. Burnside, *Theory of Groups of Finite Order*, Cambridge, 1911.

[4] W. Feit, "On the structure of Frobenius groups," *Canadian Journal of Mathematics*, vol. 9 (1957), pp. 587-596.

[5] ———, "On a class of doubly transitive permutation groups," *Illinois Journal of Mathematics* (to appear).

[6] M. Suzuki, "On the characterization of linear groups I," *Transactions of the American Mathematical Society*, vol. 92 (1959), pp. 191-204.

[7] H. Zassenhaus, *The Theory of Groups*, New York, 1958.

[8] ———, "Kennzeichnung endlicher linearen Gruppen als Permutationsgruppen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 11 (1936), pp. 17-40.

DECOMPOSITION INTO FOUR INTEGRAL SQUARES IN THE FIELDS OF $2^{\frac{1}{2}}$ AND $3^{\frac{1}{2}}$.* [1]

By Harvey Cohn.

1. **Introduction.** In a startling paper in 1928, Fritz Götzky [3] proved that every totally positive integer in the field of $5^{\frac{1}{2}}$ is representable as the sum of four integral squares at least once, by proving an exact formula analogous to the formula for the number of representations of a positive integer as the sum of four rational integral squares. Just as the latter formula was proved by Jacobi using ordinary modular functions [4], so was Götzky's formula proved by using modular functions of two complex variables.

Götzky's paper was only a partial realization of a program of Hecke [7], yet it is almost impossible to summarize the work it has influenced [13]. One might cite papers of Maass [11, 12] developing a special property related to $5^{\frac{1}{2}}$ and proving for example that *three* squares suffice instead of four, with a formula analogous to that of Gauss [4] for the number of representations of a positive integer as the sum of three squares of rational integers. Another line of approach [8, 15] was to consider in how broad a context (fields, matrices, etc.) the transformation theory of theta-functions is applicable. Our purpose is basically to show that $2^{\frac{1}{2}}$ is as "felicitous" as $5^{\frac{1}{2}}$, which seems to have been unappreciated in the past. (Indeed, $3^{\frac{1}{2}}$ is almost as good.)

Recent large scale calculations of the author indicated that the field of $2^{\frac{1}{2}}$ is also endowed with the property that every totally positive integer,

$$\mu = a + b2^{\frac{1}{2}}, \qquad a > |b|^{\frac{1}{2}} \geqq 0 \tag{1}$$

is representable as the sum of four squares if b is even. (Clearly if b is odd no such representation can occur). In fact for the extent of the experiment $(a < 147)$ even *three* squares suffice. We make no attempt to pursue this matter further here although a conjecture seems warranted [3].

In the present paper we prove the following theorem: The number of representations $A(\mu)$ of μ as the sum of four squares (with b even) in the field of $2^{\frac{1}{2}}$ is given by $B(\mu)$ where, by definition,

* Received May 1, 1959.

[1] Work supported by Research Grant G-7412 of the National Science Foundation. Presented to the American Mathematical Society, September 2, 1959.

$$(2)\qquad \begin{cases} B(\mu) = 8G(\mu), & N(\mu) \text{ odd}, \\ B(\mu) = 32G(\mu), & 4 \mid N(\mu), 8 \nmid N(\mu), \\ B(\mu) = 48G(\mu) + 6H(\mu), & 8 \mid N(\mu), \end{cases}$$

where

$$(3)\qquad \begin{cases} G(\mu) = \sum \mid N(\nu) \mid, & (\nu) \mid \mu, N(\nu) \text{ odd}, \\ H(\mu) = \sum \mid N(\nu) \mid, & (\nu) \mid \mu,\ 16 \mid N(\nu). \end{cases}$$

Here we use the symbol $(\nu) \mid \mu$ to indicate that all factors that constitute the same ideal (ν), or differ by a unit factor, are counted only once. Since b is even then $N(\mu)$ can not be even unless $4 \mid N(\mu)$. The sum $H(\mu)$ will be vacuous $(= 0)$ if $16 \nmid N(\mu)$.

The proof we present will follow the familiar pattern of Götzky's proof in that the fourth power of a theta-function will be shown to satisfy the same transformation group as an Eisenstein series, yielding the formula (2) from coefficients. The basic difference is that the transformation of the theta functions follows a somewhat different pattern in $R(2^{\frac{1}{2}})$ than in $R(5^{\frac{1}{2}})$ owing to the fact that the residue class ring mod 2 in $R(2^{\frac{1}{2}})$, (unlike the residue class ring mod 2 in $R(5^{\frac{1}{2}})$), has a zero divisor, namely $2^{\frac{1}{2}}$, yielding two "infinite" elements, $1/0$ and $1/2^{\frac{1}{2}}$. Götzky's procedure for locating the zeros of the theta-function can be simplified somewhat, requiring less knowledge of the fundamental domain that arises. Because of the detailed estimates carried out by Götzky for $R(5^{\frac{1}{2}})$, the reader can be spared the repetition of majorizing estimates.

On the basis of further numerical work, the author would readily conjecture [3] that the decomposition into four (in fact three) quadratic integral squares is likewise valid in $R(3^{\frac{1}{2}})$. Here an effort will be made to match an Eisenstein series to the corresponding theta-function with the result that the difference of the two functions is a cusp form which is *not* identically zero (as compared with the case of $R(2^{\frac{1}{2}})$ and $R(5^{\frac{1}{2}})$). The cusp form is still of some intrinsic interest since in addition to providing an approximation (in the sense that a cusp form has "small" coefficients), the cusp form provides number-theoretic identities which seem scarcely provable otherwise.

2. Notation, fundamental domain. We follow the usual custom of using Roman letters for rational integers, Greek letters for algebraic integers (which may specialize to rationals), as well as some special symbols, such as

$$(4)\qquad \begin{cases} \tau = r + si, & s > 0, \\ \tau' = r' - s'i, & s' > 0, \end{cases}$$

for the two complex variables. We use the symbols α, α', etc., to denote conjugate quadratic integers. Hence in the natural manner we generalize the concepts of norm N and trace S, e. g.,

$$(5)\quad \begin{cases} N((\alpha\tau+\beta)/(\gamma\tau+\delta)) = [(\alpha\tau+\beta)/(\gamma\tau+\delta)]\cdot[(\alpha'\tau'+\beta')/(\gamma'\tau'+\delta')], \\ S(\nu\tau/D^{\frac{1}{2}}) = (\nu\tau-\nu'\tau')/D^{\frac{1}{2}}. \end{cases}$$

We consider only the fields $R(D^{\frac{1}{2}})$, where $D=5, 8, 12$ the field discriminant corresponding to $5^{\frac{1}{2}}$, $2^{\frac{1}{2}}$, $3^{\frac{1}{2}}$. These are real quadratic fields for which the Euclidean algorithm is valid. We consider the (unimodular) group $\mathfrak{G}_D$ of transformations

$$(6)\qquad \mathfrak{G}_D:\ \tau\rightarrow(\alpha\tau+\beta)/(\gamma\tau+\delta),\qquad \tau'\rightarrow(\alpha'\tau'+\beta')/(\gamma'\tau'+\delta')$$

represented by the single matrix $\begin{pmatrix}\alpha\beta\\ \gamma\delta\end{pmatrix}$, which is to be unimodular, i. e., $\alpha\delta-\beta\gamma=\alpha'\delta'-\beta'\gamma'=1$ and of course α, α' etc., are integers in $R(D^{\frac{1}{2}})$. For convenience we shall use matrix form with the understanding that the negative of a matrix is identified with the matrix.

THEOREM 1. *The group $\mathfrak{G}_D$ is generated by the matrices*

$$\mathfrak{S}_1=\begin{pmatrix}1&1\\0&1\end{pmatrix},\qquad \mathfrak{T}=\begin{pmatrix}0&-1\\1&0\end{pmatrix},\qquad \mathfrak{U}=\begin{pmatrix}\epsilon&0\\0&\epsilon^{-1}\end{pmatrix},$$

and $\mathfrak{S}_\epsilon=\begin{pmatrix}1&\epsilon\\0&1\end{pmatrix}$, *where the ϵ are fundamental units:*

$$(7)\qquad \begin{cases}\epsilon=\frac{1}{2}(1+5^{\frac{1}{2}}) & \text{for } D=5,\\ \epsilon=1+2^{\frac{1}{2}} & \text{for } D=8,\\ \epsilon=2+3^{\frac{1}{2}} & \text{for } D=12.\end{cases}$$

Götzky restricted his proof to $D=5$, which has the distinction that one can do away with $\mathfrak{S}_\epsilon$, since $\mathfrak{S}_\epsilon=\mathfrak{S}_1^{-1}\mathfrak{U}\mathfrak{S}_1\mathfrak{U}^{-1}$. Otherwise the proof is no different since it is based wholly on the Euclidean algorithm: If δ and γ $(\neq 0)$ are two integers of $R(D^{\frac{1}{2}})$, an algebraic integer ν evists such that $|N(\gamma\nu-\delta)|<|N(\gamma)|$.

Now under the equivalence classes set up in $\mathfrak{G}_D$ the four-dimensional real (or two-dimensional complex) manifold of (τ,τ') is generated by operations on a *fundamental domain.* Using Götzky's construction for $R(5^{\frac{1}{2}})$, we can say generally, the *fundamental domain* is a connected set in which the following inequalities are valid:

$$(8)\qquad \begin{cases}-\frac{1}{2}\leqq S(r/D^{\frac{1}{2}})<\frac{1}{2}\\ -\frac{1}{2}\leqq S(\epsilon r/D^{\frac{1}{2}})<\frac{1}{2}\end{cases}$$

$$(9)\qquad \epsilon'^2<s/s'<\epsilon^2$$

8

The inequalities (8) are ensured by the presence in the group $\mathfrak{G}$ of all translations $\mathfrak{S}_1^a \mathfrak{S}_\epsilon^b$ or $\tau \to \tau + (a + b\epsilon)$. Here, in each case $\epsilon - \epsilon' = D^{\frac{1}{2}}$. The inequalities (9) are a result of the repeated application of $\mathfrak{U}$. These inqualities are not sufficient to determine the "floor" of the fundamental domain. For our purpose it is sufficient to use the estimate

$$ss' \geqq 2/D \tag{10}$$

which together with (9) yields lower bounds for s and s', namely

$$\begin{cases} s \geqq |\epsilon'| (2/D)^{\frac{1}{2}} \\ s' \geqq |\epsilon'| (2/D)^{\frac{1}{2}}. \end{cases} \tag{10a}$$

Applying the transformations $\mathfrak{S}_1$, $\mathfrak{S}_\epsilon$, and $\mathfrak{U}$ to a set of points satisfying (10a), we can (in a simple way) obtain, with Götzky, a fundamental domain lying inside the region determined by inequalities (8), (9), and (10a). We recapitulate the most important part of our result, to obtain the following:

THEOREM 2. *Every pair of complex numbers* (τ, τ') *is equivalent under the group* $\mathfrak{G}_D$ *to a pair for which the imaginary parts are bounded away from zero, and the fundamental domain lies in the region defined by* (8), (9), *and* (10a).

The most difficult part of the proof is estimate (10) which we shall prove under the assumption that the real quadratic field has unique factorization. If we specialize the original method of Blumenthal [1], we know only $ss' > 1/4D$.

LEMMA. *If* ω_1, ω_2, *and* ω_1', ω_2' *are two pairs of complex numbers which are not proportional, e.g., if the "areas" defined by*

$$\Delta = |\operatorname{Re}\omega_1 \operatorname{Im}\omega_2 - \operatorname{Im}\omega_1 \operatorname{Re}\omega_2|$$

and the corresponding Δ' *are* $\neq 0$, *then a pair of algebraic integers* (γ, δ) *exists for which the function*

$$T = |\gamma\omega_1 + \delta\omega_2|^2 + |\gamma'\omega_1' + \delta'\omega_2'|^2 \tag{11}$$

satisfies the relation

$$0 < T \leqq (2D\Delta\Delta')^{\frac{1}{2}}. \tag{12}$$

Proof. Use the integral coordinates x, y, z, w

$$\begin{cases} \gamma = x + y\epsilon, & \gamma' = x + y\epsilon' \\ \delta = z + w\epsilon, & \delta' = z + w\epsilon', \end{cases} \tag{13}$$

and T becomes a positive definite quarternary form in x, y, z, w. Writing

$$\begin{cases} \gamma\omega_1 + \delta\omega_2 = R + iS \\ \gamma'\omega_1' + \delta'\omega_2' = R' + iS', \end{cases} \tag{14}$$

we can say $T = R^2 + S^2 + R'^2 + S'^2$; but

$$\begin{aligned} &| \partial(x,y,z,w)/\partial(R,S,R',S') | \\ &\qquad = |[\partial(x,y,z,w)/\partial(\gamma,\delta,\gamma',\delta')] \cdot [\partial(\gamma,\delta,\gamma',\delta')/\partial(R,S,R',S')] \\ &\qquad = [1/(\epsilon - \epsilon')^2] \cdot [1/\Delta\Delta'], \end{aligned}$$

and thus the determinant for the form $T = T(x,y,z,w)$ is $D^2N(\Delta)^2$. Hence using the constant $2^{\frac{1}{2}}$ supplied by Korkine and Zolatoreff [9], we obtain the relation (12), proving the lemma.

To achieve the theorem, note that $T \geqq | 2N(\gamma\omega_1 + \delta\omega_2)|$, for all (γ,δ). The same inequality is even stronger if we remove the common factor (if any) of (γ,δ), thus rendering in new symbols $(\gamma,\delta) = 1$, with

$$(2DN(\Delta))^{\frac{1}{2}} \geqq | 2N(\gamma\omega_1 + \delta\omega_2)|. \tag{15a}$$

If we let $\omega_1 = 1$, $\omega_1' = 1$, $\omega_2 = \tau$, $\omega_2' = \tau'$, we find $N(\Delta) = ss'$, and if we find the pair of integers α, β for which $\alpha\delta - \beta\gamma = 1$, then

$$| N(\Delta/(\gamma\omega_1 + \delta\omega_2))| = | N\{\mathrm{Im}((\alpha\tau + \beta)/(\gamma\tau + \delta))\}|, \tag{15b}$$

which makes relation (10) a consequence of (15a) for the equivalent τ defined by the unimodular transformation written in (15b). Q. E. D.

In this proof we assumed only the unique factorization properties in order to make $\alpha\delta - \beta\gamma = 1$. On the basis of the bounds deduced here, we can show this fundamental result:

THEOREM 3. *Any function of two complex variables which is invariant under the group $\mathfrak{G}_D$ and which has no finite singularities is a constant.*

Götzky's proof applies admirably, thanks to Theorem 2. A more general proof appears in Maass [10] and Koecher [8a], but our results are more easily seen as extensions of Götzky's proof than as specializations of the latter proofs. (Incidentally, a weaker theorem requiring boundedness at infinity would have caused no difficulty in our present application!) Finally, following Götzky, we apply symmetric functions to prove an extension:

THEOREM 4. *A finite set of functions of two complex variables which are permuted under the group $\mathfrak{G}_D$ and which have no finite singularities are constants.*

One noteworthy feature of Götzky's proof is the use of the fact that from the translation group operations $\mathfrak{S}_1$ and $\mathfrak{S}_\epsilon$, the modular function has an expansion

$$F(\tau,\tau') = A_0 + \sum A_\nu \exp[\pi i S(\nu\tau/D^{\frac{1}{2}})], \tag{16}$$

where ν is restricted to totally positive integers. In some cases it will happen that $A_\nu = A_{\nu'}$ which we describe as a *symmetric* function, but we do not make this a requirement on $F(\tau,\tau')$.

3. Theta-functions. We consider, initially, an arbitrary real quadratic field of discriminant D and fundamental unit ϵ. We define the theta-function of τ and τ' of type ρ, σ as

$$\Theta(\rho,\sigma;\tau) = \sum_\nu \exp(\pi i) S\{(\nu\sigma + (\nu+\rho/2)^2\tau)/D^{\frac{1}{2}}\}; \quad \operatorname{Im}\tau > 0, \operatorname{Im}\tau' < 0; \tag{17}$$

where ν runs over all integers of $R(D^{\frac{1}{2}})$. The significance of this function lies in the fact that

$$\Theta^4(0,0;\tau) = 1 + \sum_{\mu \gg 0} A(\mu)\exp[\pi i S\{\mu\tau/D^{\frac{1}{2}}\}], \tag{18}$$

where $\mu \gg 0$ denotes the quadratic integers that are totally positive, (i. e., for which $\mu > 0$, $\mu' > 0$), and $A(\mu)$ denotes the number of representations of μ as the sum of four squares. Now, unfortunately, $\Theta^4(0,0;\tau)$ is not a modular form under group $\mathfrak{G}_D$ but it is permuted with the set $\Theta^4(\rho,\sigma;\tau)$, which is a finite set since the values of $\Theta(\rho,\sigma;\tau)$ depend only on the residue class of ρ or $\sigma \bmod 2$. The very formal estimates of Götzky for $D = 5$ establish the absolute, uniform convergence of the series (17) for the fundamental domain.

The transformations of (17) are next considered for the generators of $\mathfrak{G}_D$. First of all, trivially,

$$\mathfrak{U}: \Theta(\rho,\sigma;\epsilon^2\tau) = \Theta(\rho\epsilon,\sigma\epsilon^{-1};\tau). \tag{19a}$$

Next, using the famous Poisson-Lipschitz formula (31), below, we find

$$T: \Theta(\rho,\sigma;1/\tau) = \exp[-\tfrac{1}{2}\pi i S(\rho\sigma/D^{\frac{1}{2}})]\Theta(-\sigma,\rho;\tau)N(\tau)^{\frac{1}{2}}. \tag{19b}$$

Here $N(\tau)^{\frac{1}{2}} = (\tau\tau')^{\frac{1}{2}}$ is chosen as the positive branch when τ,τ' are purely imaginary. To obtain the effect of $\mathfrak{S}_1$ and $\mathfrak{S}_\epsilon$ we must specialize the information further. If we assume D is even,

$$\mathfrak{S}_1: \Theta(\rho,\sigma;\tau+1) = \exp[\tfrac{1}{4}\pi i S(\rho^2/D^{\frac{1}{2}})]\Theta(\rho,\sigma+\rho;\tau). \tag{19c}$$

Likewise

(19d) $\quad \mathfrak{S}_\epsilon\colon \Theta(\rho, \sigma; \tau + \epsilon) = \exp[\tfrac{1}{4}\pi i S(\rho^2\epsilon/D^{\frac{1}{2}})]\Theta(\rho, \sigma + \epsilon\rho + \epsilon + 1; \tau)$

under the further assumption (embracing both $D = 8$, $D = 12$),

$$\begin{cases} \epsilon = g + m^{\frac{1}{2}} \\ D = 4m. \end{cases} \tag{20}$$

We do not need to use all possible Θ functions. First let us consider η a representative of the residue class given by

$$\begin{cases} \eta^2 \equiv 0 \mod 2 \\ \eta \not\equiv 0 \mod 2. \end{cases} \tag{21a}$$

Such an η is, by definition, $\equiv 1 + \epsilon \pmod 2$. In specific cases

$$\begin{cases} \eta = 2^{\frac{1}{2}} \text{ for } D = 8 \\ \eta = 1 + 3^{\frac{1}{2}} \text{ for } D = 12. \end{cases} \tag{21b}$$

Now we can restrict ourselves to these four types, basically reducing "mod 2" to "mod η";

$$\Theta_{d,c}(\tau) = \Theta(d\eta, c\eta; \tau);\ c, d \equiv 0, 1 \pmod 2. \tag{22}$$

Then, referring to the generators of $\mathfrak{G}_D$, we find

$$\mathfrak{U}\colon \Theta^4_{d,c}(\epsilon^2\tau) = \Theta^4_{d,c}(\tau), \tag{23a}$$

$$\mathfrak{T}\colon \Theta^4_{d,c}(-1/\tau) = \Theta^4_{c,d}(\tau)N(\tau)^2, \tag{23b}$$

$$\mathfrak{S}_1\colon \Theta^4_{d,c}(\tau + 1) = \Theta^4_{d,c+d}(\tau), \tag{23c}$$

$$\mathfrak{S}_\epsilon\colon \Theta^4_{d,c}(\tau + \epsilon) = \Theta^4_{d,c+d+1}(\tau). \tag{23d}$$

This system of equations, together with the fact that each $\Theta_{d,c}(\tau)$ is bounded as $\operatorname{Im}\tau$, $-\operatorname{Im}\tau' \to \infty$ and is free of singularities, defines a set of *modular forms* of dimension -2 belonging to the $\mathfrak{G}_D$.

More generally if $\mathfrak{S}$ is a transformation of $\mathfrak{G}_D$, we would find

$$\mathfrak{S}\colon \Theta^4_{d,c}(\mathfrak{S}\tau) = \Theta^4_{d',c'}(\tau)N(\gamma\tau + \delta)^2, \tag{24}$$

where $\mathfrak{S}\tau = (\alpha\tau + \beta)/(\gamma\tau + \delta)$ (in lowest terms) and d', c' are linear inhomogeneous combinations of d and c modulo 2 (which need not be specified more generally). A set of four modular forms with these indices, subject to these transformations is said to *transform* like $\Theta^4_{0,0}(\tau)$. The transformations preserving the indices $d' = d = c' = c = 0$ form a sub-group $\mathfrak{H}_D$ of index 4 in $\mathfrak{G}_D$ containing $\mathfrak{S}_1$, $\mathfrak{T}$, and $\mathfrak{U}$, but not $\mathfrak{S}_\epsilon$. (See Section 6, below). This set of conjugates is then said to *belong to* $\mathfrak{H}_D$.

The conjugates are achieved by the following operations on $\Theta^4_{0,0}(\tau)$:

(24a) $$\mathfrak{S}_\epsilon\colon \Theta^4_{0,0}(\tau+\epsilon)=\Theta^4_{0,1}(\tau),$$

(24b) $$\mathfrak{S}_\epsilon\mathfrak{T}\colon \Theta^4_{0,1}(-1/\tau)=\Theta^4_{1,0}(\tau)N(\tau)^2,$$

(24c) $$\mathfrak{S}_\epsilon\mathfrak{T}\mathfrak{S}_1\colon \Theta^4_{1,0}(\tau+1)=\Theta^4_{1,1}(\tau).$$

It can be seen that as $\operatorname{Im}\tau$, $-\operatorname{Im}\tau'\to\infty$

(25a) $$\Theta^4_{0,0}(\tau)\to 1,$$

(25b) $$\Theta^4_{0,1}(\tau)\to 1,$$

(25c) $$\Theta^4_{1,0}(\tau)\to 0,$$

(25d) $$\Theta^4_{1,1}(\tau)\to 0.$$

This can be done by writing out terms and noting uniform convergence.

The *fourth* power of $\Theta_{d,c}(\tau)$ is incidentally seen to be the least for which all the factors of transformations (19b, c, d) become unity. Note that the ρ^2 occurring in (19c, d) is always divisible by $(2)=(\eta^2)$, in the set-up of equation (22).

The reader will note that the situation here is considerably simpler than that of Götzky where $D=5$ and $[\mathfrak{G}_D\colon \mathfrak{H}_D]=10$. There 10 theta-functions occur! (The condition is even more difficult for a general field [8]).

4. Eisenstein series. Following Hecke's method we define the Eisenstein series in τ and τ', abbreviated

(26a) $$A_\kappa(\tau)=\sum_{(\nu),\mu}[\nu\tau+\mu], \qquad \nu\kappa\equiv\mu \bmod 2.$$

where we use the symbol

(26b) $$[\nu\tau+\mu]=N\{(\nu\tau+\mu)\,|\,\nu\tau+\mu\,|^k\}^{-2}, \qquad k>0,$$

and in particular

(26c) $$[\tau]=(\tau\tau')^{-2}\,|\,\tau\tau'\,|^{-2k}, \text{ etc.}$$

and the indices of summation "$(\nu),\mu$" mean, as before, that only the ideal (ν) is represented, (associated numbers are not repeated), while μ runs over all quadratic integers satisfying the indicated congruence. Clearly $A_\kappa(\tau)$ depends only on the residue class of $\kappa \bmod 2$.

Now Götzky's paper has a complete discussion of majorants (requiring only Theorem 2 and applicable without effort to the present case). The

manipulations themselves are more distinctive, however, and will need to be summarized here. We introduce the symbol, for each D,

(26d) $$\zeta_k = \sum_{(\nu)} 1/|N(\nu)|^{2+2k}.$$

Then if $\kappa \not\equiv 0 \bmod \eta$,

$$\begin{aligned} A_\kappa(\tau) &= \sum_{\substack{\mu=0 \\ (\nu)=(2\sigma)}} [\nu\tau+\mu] + \sum_{\substack{(\nu),\mu\neq 0 \\ \nu\kappa\equiv\mu \bmod 2}} [\nu\tau+\mu] \\ &= \zeta_k/16^{1+k}[\tau] + \sum_{\substack{(\nu),\mu\neq 0 \\ \nu\equiv\mu\kappa^{-1} \bmod 2}} [\nu\tau+\mu] \\ &= \zeta_k/16^{1+k}[\tau] + \sum_{\substack{\nu,(\mu)\neq 0 \\ \nu\equiv\mu\kappa^{-1} \bmod 2}} [\mu(-1/\tau)+\nu]\cdot[\tau]^{-1}. \end{aligned}$$

Thus

(27a) $$\{A_\kappa(\tau)+\zeta_k/16^{1+k}\} = [\tau]^{-1}\{A_{\kappa^{-1}}(-1/\tau)+\zeta_k/16^{1+k}\},\ \kappa\not\equiv 0 \bmod \eta.$$

By similar manipulations, we take $\kappa \equiv 0 \bmod 2$ and find

(27b) $$\{A_0(2\tau)+\zeta_k/16^{1+k}\} = [\tau]^{-1}\{A_0(-2/\tau)+\zeta_k/16^{1+k}\}.$$

Now, finally, if $\eta | \kappa$ and $2 \nmid \kappa$, (noting $\eta \equiv \eta' \bmod 2$), we have

(27c) $$\{A_\eta(\eta\tau)+\zeta_k/16^{1+k}\} = [\tau]^{-1}\{A_\eta(-\eta'/\tau)+\zeta_k/16^{1+k}\}.$$

In the last formula it will be convenient to have

(28) $$\eta\eta' = -2$$

as is clearly the case with formula (21b).

In addition to the indicated "inversion" type formulae we have the translation formula

(29) $$A_\kappa(\tau+\rho) = A_{\kappa+\rho}(\tau),$$

and the unit formula (noting $\epsilon^2 \equiv 1 \bmod 2$),

(30) $$A_\kappa(\epsilon^2\tau) = A_\kappa(\tau).$$

The formulas (27abc), (29), and (30) take care of transformations $\mathfrak{T}$, $\mathfrak{S}_1{}^a\mathfrak{S}_\epsilon{}^b$ and $\mathfrak{U}$ respectively (where $\rho = a + b\epsilon$).

5. Expansion into Fourier series. Now the translation formula (29) assures us that a double Fourier Series for $A_\kappa(\tau)$ can be found since $A_\kappa(\tau+2\rho) = A_\kappa(\tau)$. To simplify the procedure of Hecke and Götzky somewhat we can use the following result:

POISSON-LIPSCHITZ LEMMA. *Assuming absolute convergence of all indicated processes:*

$$(31)\qquad \sum_{\rho} Q(\rho,\rho') = (1/D^{\frac{1}{2}}) \sum_{\mu} \int\int_{-\infty}^{\infty} Q(X,X')\exp[-2\pi i S(X\mu/D^{\frac{1}{2}})]\,dX\,dX',$$

where X, X' are real variables and μ is a general integer of $R(D^{\frac{1}{2}})$.

Hence we can write, setting $\mu = 2\rho + \nu\kappa$,

$$(32)\qquad A_{\kappa}(\tau) = \sum_{(\nu)} (1/16^{1+k}) \sum_{\rho} Q(\rho,\rho'),$$

where

$$(33)\qquad Q(\rho,\rho') = [(\nu\tau + \nu\kappa)/2 + \rho].$$

According to estimates parallel to those in Götzky, the limit as $k \to 0$ can be performed under the integral sign in (31) and the method of residues can be brought into play. Each double integral is of the type

$$(34)\qquad \int_{-\infty}^{\infty} \{(\nu\tau + \nu\kappa)/2 + X\}^{-2} \exp(-2\pi i X\mu/D^{\frac{1}{2}})\,dX \times \int_{-\infty}^{\infty} \{(\nu'\tau' + \nu'\kappa')/2 + X'\}^{-2} \exp(+2\pi i X'\mu'/D^{\frac{1}{2}})\,dX'$$

It is easily seen that the first factor is zero unless $\nu\mu > 0$ since $\operatorname{Im}\tau > 0$ while the second factor is zero unless $\nu'\mu' > 0$ since $\operatorname{Im}\tau' < 0$, and in each factor the contour will be deformed to infinity. The residue of $(\exp Az)/(z+B)^2$ is easily seen to be $A\exp(-AB)$ at the pole $z = -B$.

Thus combining all the foregoing results and letting $k \to 0$, we see that for the four residue classes κ (mod 2)

$$(35)\qquad \lim_{k\to 0} A_{\kappa}(\tau) = (\pi^4/D^{3/2}) F_{\kappa}(\tau) \qquad (\kappa \equiv 0, 1, \eta, 1+\eta),$$

where

$$(36)\qquad F_{\kappa}(\tau) = \sum_{\substack{(\nu),\mu \\ \nu\mu \gg 0}} \exp[\pi i S(\nu\kappa\mu/D^{\frac{1}{2}})] \exp[\pi i S(\nu\mu\tau/D^{\frac{1}{2}})] \,|N(\mu)|.$$

Hence if we introduce the additional constant

$$(37)\qquad C_D = (D^{3/2}/16\pi^4) \sum_{(\nu)} 1/N(\nu)^2,$$

we find that the new functions

$$(38)\qquad \psi_{\kappa}(\tau) = C_D + F_{\kappa}(\tau)$$

will inherit valuable invariance properties from formulae (27abc), (29), and (30); (see the next section).

We first evaluate C_D. By the well-known algebraic zeta-function procedures if $\mathfrak{p}$ denotes the prime ideals of $R(D^{\frac{1}{2}})$,

$$(39)\qquad C_D = (D^{3/2}/16\pi^4)\prod_{\mathfrak{p}}(1-N(\mathfrak{p})^{-2})^{-1}$$
$$= (D^{3/2}/16\pi^4)\prod_{p}(1-p^{-2})^{-1}\cdot\prod_{p}(1-(D/p)p^{-2})^{-1},$$

where (D/p) is the Jacobi-symbol and p runs over rational primes. Since for the ordinary zeta-function $\zeta(2)=\pi^2/6$,

$$(40)\qquad C_D = (D^{3/2}/96\pi^2)\sum_{q=1}^{\infty}(D/q)q^{-2}.$$

But since $(D/q)=(D/D-q)$, we can let Q run through half the residue classes mod D. Thus, from the familiar expansion,

$$(41)\qquad \pi^2/(\sin\pi Q/D)^2 = \sum_{-\infty}^{\infty}1/(Q/D-a)^2$$
$$= D^2\{\sum_{\substack{q>0\\ q\equiv Q \bmod D}} 1/q^2 + \sum_{\substack{q>0\\ q\equiv -Q \bmod D}} 1/q^2\}.$$

Thus, finally, using the symmetry on Q,

$$(42)\qquad C_D = (1/192D^{\frac{1}{2}})\sum_{Q=1}^{D}(D/Q)/\sin^2\pi Q/D.$$

In particular $C_8=1/48$ and $C_{12}=1/24$.

6. New modular forms. We now define in addition to $\psi_\kappa(\tau)$ of formula (38), the additional functions

$$(43)\qquad \begin{cases}\psi_\infty(\tau)=16\psi_0(4\tau)\\ \psi_{1/\eta}(\tau)=4\psi_\eta(2\tau)\end{cases}$$

yielding $\psi_\lambda(\tau)$ for a set of λ consisting of the finite elements (mod 2)

$$(44a)\qquad \kappa = 0, 1, \eta, \epsilon,$$

and the "infinite" elements ∞, $1/\eta$ to make a complete set (mod 2)

$$(44b)\qquad \lambda = 0, 1, \eta, \epsilon, 1/\eta, \infty.$$

We then have the following rules by virtue of relations (27abc), (29), (43), using κ $(=a+b\epsilon)$ to denote a finite index and λ to denote a general index:

$$(45a)\qquad \psi_\lambda(\tau+\kappa)=\psi_{\lambda+\kappa}(\tau) \qquad (\mathfrak{S}_1^{a}\mathfrak{S}_\epsilon^{b}=\mathfrak{S}_\kappa),$$

$$(45b)\qquad \psi_\lambda(\epsilon^2\tau)=\psi_\lambda(\tau) \qquad (\mathfrak{U}),$$

$$(45c)\qquad \psi_\lambda(-1/\tau)=\psi_{1/\lambda}(\tau)N(\tau)^2 \qquad (\mathfrak{T}).$$

We can verify furthermore that $\psi_0(2\tau)$ is itself invariant under all generating transformations of $\mathfrak{G}_D$.

More generally we can write for any transformation $\mathfrak{S}$ of $\mathfrak{G}_D$

$$\psi_\lambda(\mathfrak{S}\tau) = \psi_{\mathfrak{S}\lambda}(\tau) N(\gamma\tau + \delta)^2, \tag{46}$$

where $\mathfrak{S}\tau = (\alpha\tau + \beta)/(\gamma\tau + \delta)$ in "lowest terms" (as is usual). Here the symbol $\mathfrak{S}\lambda$ always has meaning since each λ, finite or infinite, can be expressed as the ratio of two finite element not both divisible by η, $\lambda = \kappa_1/\kappa_2$. Then $\mathfrak{S}\lambda$ can be expressed in terms of homogeneous transformations on κ_1 and κ_2 in which numerator and denominator are never both $\equiv 0 \bmod \eta$ by the unimodular nature of the determinant. The equation (46) definies a modular form of dimension -2.

We therefore have a system of six functions $\psi_\lambda(\tau)$ permuted by the transformations of $\mathfrak{G}_D$. If we wish to match $\Theta^4{}_{0,0}(\tau)$, we must first find a linear combination invariant under the transformations of $\mathfrak{S}_1$, $\mathfrak{T}$, $\mathfrak{U}$ (but not necessarily $\mathfrak{S}_\epsilon$, as shown in equations (23a-d)). Since the sets of values $\lambda = 0, 1, \infty$ and $\lambda = \eta, \epsilon, 1/\eta$ are permuted under these transformations, it follows that the functions

$$\Omega_1(\tau) = \psi_0(\tau) + \psi_1(\tau) + \psi_\infty(\tau), \tag{47a}$$

$$\Omega_2(\tau) = \psi_\eta(\tau) + \psi_\epsilon(\tau) + \psi_{1/\eta}(\tau), \tag{47b}$$

$$\Omega_3(\tau) = \psi_0(\tau), \tag{47c}$$

are modular forms of dimension -2 which are invariant under the transformations $\mathfrak{S}_1$, $\mathfrak{T}$ and $\mathfrak{U}$.

We can now verify the forms are not independent but

$$\Omega_1(\tau) + \Omega_2(\tau) = 24\Omega_3(\tau). \tag{48}$$

This can be done by the actual examination of the Fourier Series of formula (36). For we must show

$$\psi_0(\tau) + \psi_1(\tau) + \psi_\epsilon(\tau) + \psi_\eta(\tau) + 4\psi_\eta(2\tau) + 16\psi_0(4\tau) = 24\psi_0(2\tau).$$

The constant terms are easily taken into account. Then we consider

$$\psi^*(\tau) = \psi_0(\tau) + \psi_1(\tau) + \psi_\epsilon(\tau) + \psi_\eta(\tau) = \sum_\kappa \psi_\kappa(\tau).$$

We find

$$\psi^*(\tau) = 4C_D + \sum_{\substack{(\nu),\mu \\ \nu\mu \gg 0}} \{\sum_\kappa \exp[\pi i S(\nu\mu\kappa/D^{\frac{1}{2}})]\} \exp[\pi i S(\nu\mu\tau/D^{\frac{1}{2}})] |N(\mu)|. \tag{49}$$

It is easily seen that the "$\{\cdot\cdot\cdot\}$" expression is 0 unless $\nu\mu$ is even in which case it is 4. In fact if $\nu\mu = a + b\epsilon$, $\{\cdot\cdot\cdot\} = [1 + (-1)^a][1 + (-1)^b]$. Hence identity (48) reduces to the new identity

$$\psi^*(\tau) + 4\psi_\eta(2\tau) + 16\psi_0(4\tau) = 24\psi_0(2\tau). \tag{50}$$

Now we let $2\lambda = \mu\nu$ in $\psi^*(\tau)$ and $\lambda = \mu\nu$ in the other terms and we find that the identity (50) is the following (easily demonstrable) results when referred to the coefficient of $\exp[2\pi i S(\lambda\tau/D^{\frac{1}{2}})]$.

If $\eta\nmid\lambda;\ 4\sum_{(\nu)|2\lambda}|N(\nu)| - 4\sum_{(\nu)|\lambda}|N(\nu)| + 0 = 24\sum_{(\nu)|\lambda}|N(\nu)|.$

If $\eta|\lambda, 2\nmid\lambda;\ 4\sum_{(\nu)|2\lambda}|N(\nu)| + 4\sum_{(\nu)|\lambda}|N(\nu)| + 0 = 24\sum_{(\nu)|\lambda}|N(\nu)|.$

If $2|\lambda;\ 4\sum_{(\nu)|2\lambda}|N(\nu)| + 4\sum_{(\nu)|\lambda}|N(\nu)| + 16\sum_{(\nu)|2\lambda}|N(\nu)| = 24\sum_{(\nu)|\lambda}|N(\nu)|.$

It will turn out on the basis of later results that $\Omega_1(\tau)$ and $\Omega_3(\tau)$ form a basis of modular forms of dimension -2 belonging to $\mathfrak{H}_D$ the group of $\Theta^4{}_{0,0}(\tau)$, (which will yield the desired expression for $\Theta^4{}_{0,0}(\tau)$), when $D = 8$. This is not true when $D = 12$. One special linear combination of $\Omega_1(\tau)$ and $\Omega_3(\tau)$ is of special interest in any case, namely

$$\Omega_{0,0}(\tau) = (12C_D)^{-1}[\psi_0(\tau) + \psi_1(\tau) + \psi_\infty(\tau) - 6\psi_0(2\tau)]. \tag{51a}$$

Its importance becomes apparent if we define Ω_{dc} to transform like $\Theta^4{}_{d,c}$ as in equation (24abc):

$$\Omega_{0,1}(\tau) = (12C_D)^{-1}[\psi_\epsilon(\tau) + \psi_\eta(\tau) + \psi_\infty(\tau) - 6\psi_0(2\tau)], \tag{51b}$$

$$\Omega_{1,0}(\tau) = (12C_D)^{-1}[\psi_\epsilon(\tau) + \psi_{1/\eta}(\tau) + \psi_0(\tau) - 6\psi_0(2\tau)], \tag{51c}$$

$$\Omega_{1,1}(\tau) = (12C_D)^{-1}[\psi_\eta(\tau) + \psi_{1/\eta}(\tau) + \psi_1(\tau) - 6\psi_0(2\tau)]. \tag{51d}$$

We then can state that $\Omega_{0,0}(\tau)$ is the only linear combination whose "constant terms" (as in 25a-d)), for all conjugates $\Omega_{d,c}(\tau)$, match those for $\Theta^4{}_{d,c}(\tau)$. We can say more concisely that all cases are comprised in this one formula:

$$\Omega_{d,c}(\tau) = (12C_D)^{-1}[\psi_{1+(d+c)\eta}(\tau) + \psi_{1/(d\eta)}(\tau) + \psi_{c\eta}(\tau) - 6\psi_0(2\tau)]. \tag{52}$$

Thus we have obtained a set of functions $\Omega_{d,c}(\tau)$ which behave at ∞ and which transform like $\Theta^4{}_{d,c}(\tau)$ under the generating transformations (23a-d) of $\mathfrak{G}_D$.

Once we know that the sub-group $\mathfrak{H}_D$ is equally well determined by $\Omega_{0,0}(\tau)$, we can conclude that an arbitrary transformation $\mathfrak{S}\colon \tau \to (\alpha\tau + \beta)/(\gamma\tau + \delta)$ of $\mathfrak{G}_D$ belongs to $\mathfrak{H}_D$ if and only if $\alpha, \beta, \gamma, \delta$ are congruent to rational integers

modulo 2. From equation (52), $\mathfrak{H}_D$ must depend on the arithmetic of the indices of $\psi_\lambda(\tau)$ and $\alpha, \beta, \gamma, \delta$ modulo 2, and the rest is a matter of inspection. In particular $\mathfrak{H}_D$ contains $\mathfrak{S}_\epsilon^2$. We leave open the question of whether or not $\mathfrak{H}_D$ is generated by $\mathfrak{S}_1$, $\mathfrak{T}$ and $\mathfrak{U}$ for $D = 8$ and 12.

7. Zero manifold for $2^{1/2}$. We next establish the result that within the fundamental domain, if $N(\epsilon) = -1$, the manifold

$$\epsilon\tau = \epsilon'\tau' \tag{53}$$

provides the only points at which $\Theta_{1,1}(\tau)$ vanishes, while the other three $\Theta_{d,c}(\tau)$ are non-vanishing at finite points. We shall prove this result, essentially as Götzky proved for $D = 5$, by an appropriate rearrangement of terms. Our proof will be specialized for convenience to $D = 8$ but the method will be clearly more general.

We first of all establish the result that in the fundamental domain if $\operatorname{Im}\tau$ and $-\operatorname{Im}\tau'$ each approach infinity (under the restrictions (9)), then the zeros (53) are ultimately those described in the statement. No difficulty is present by $\Theta_{0,c}$ according to statement (25a). We introduce the new notation

$$h_\epsilon = \exp[\pi i S(\epsilon\tau/8^{1/2})], \qquad h_{\epsilon'} = \exp[\pi i S(\epsilon'\tau/8^{1/2})] \tag{54}$$

and we find with $\nu = a + b2^{1/2}$,

$$\Theta_{1,c}(\tau) = \sum_{a,b} (-)^{ac} h_{\epsilon'}^{\frac12(-a+b+\frac12)^2+\frac12(b+\frac12)^2} h_\epsilon^{\frac12(a+b+\frac12)^2+\frac12(a+\frac12)^2}. \tag{55a}$$

We write $a \to 2b - a + 1$ in order to take advantage of a natural symmetry. This changes the exponent on h_ϵ only:

$$\Theta_{1,c}(\tau) = \sum_{a,b} (-)^{ac}(-)^{c} h_{\epsilon'}^{\frac12(-a+b+\frac12)^2+\frac12(b+\frac12)^2} h_\epsilon^{\frac12(3b-a+\frac32)^2+\frac12(b+\frac12)^2}. \tag{55b}$$

Averaging, we find

$$\begin{aligned} &2\Theta_{1,c}(\tau)/(1+(-)^c h_\epsilon) \\ &\quad = \sum_{a,b} (-)^{ac} h_{\epsilon'}^{\frac12(-a+b+\frac12)^2+\frac12(b+\frac12)^2} h_\epsilon^{\frac12(a+b+\frac12)^2+\frac12(b+\frac12)^2} \cdot f_{a,b}, \end{aligned} \tag{56}$$

where

$$f_{a,b} = [1 + (-)^c h_\epsilon^{(2(b-a)+1)(2b+1)}]/(1+(-)^c h_\epsilon). \tag{57}$$

We find that the summation in (56) is similar to that of $\Theta_{1,\dot{c}}$ except for the factor $f_{a,b}$. But since $|h_\epsilon| \leqq 1$ by condition (9),

$$|f_{a,b}| \leqq |[2(b-a)+1][2b+1]|,$$

a factor too small to upset the uniform convergence of the theta-function. Hence the uniform convergence transfers to equation (56). Even more remarkable is the fact that by virtue of the factor $f_{a,b}$ the series in (56) has a symmetry property (which the series in (55a) does not possess), namely:

$$\sum_{a,b} = \sum_{\substack{2b+1>0 \\ 2b-2a+1>0}} + \sum_{\substack{2b+1>0 \\ 2b-2a+1<0}} + \sum_{\substack{2b+1<0 \\ 2b-2a+1>0}} + \sum_{\substack{2b+1<0 \\ 2b-2a+1<0}}, \tag{58}$$

while all four partial sums are *equal*. Hence the new series

$$\begin{aligned} &\Theta_{1,c}(h_\epsilon h_{\epsilon'})^{-\frac14}/2\,(1+(-)^c h_\epsilon) \\ &\quad = \sum_{\substack{2b+1>0 \\ 2b-2a+1>0}} (-)^{ac} h_{\epsilon'}{}^{[\frac12(-a+b+\frac12)^2+\frac12(b+\frac12)^2-\frac14]}\, h_\epsilon{}^{[\frac12(a+b+\frac12)^2+\frac12(b+\frac12)^2-\frac14]} f_{a,b} \end{aligned} \tag{59}$$

begins with the constant term $f_{0,0}$ $(=2)$, which must be the limit of the series as $\operatorname{Im}\tau$, $-\operatorname{Im}\tau'\to\infty$ in accordance with (9). The left hand side, furthermore, supplies the zeros of the theta-function as given by $1+(-)^c h_\epsilon = 0$, $h_\epsilon = 0$, $h_{\epsilon'} = 0$. The only zeros lying in the fundamental domain are easily seen to be those of (53) for $c=1$, i. e., zeros of $\Omega_{1,1}(\tau)$, at least in the limiting case.

We can then easily verify that *all zeros are now known*, and not just those for large $\operatorname{Im}\tau$, $-\operatorname{Im}\tau'$. The manifold of zeros is defined by $\Theta_{c,d}(\tau,\tau')=0$ which, taken as a mapping function of the τ-projection into the τ'-projection, is open at every point owing to the absence of singularities as well as the absence of "exceptional" points (where, say, $\partial\Theta_{c,d}(\tau,\tau')/\partial\tau\equiv 0$ in τ' when $\tau=\tau_0$). Therefore the τ and τ' projections of the zero manifold of (say) $\Theta_{0,c}(\tau)$ must extend to infinity if there is even one such point in the fundamental domain. Points near infinity are excluded, however, negating the existence of any zeros of $\Theta_{0,c}(\tau)$. We likewise can take care of the function in (59), etc.

8. Behavior on the zero manifold. We next consider a general modular form of dimension -2 which transforms like $\Theta^4{}_{0,0}(\tau)$. We call the modular form $\Xi(\tau)=\Xi_{0,0}(\tau)$ and define $\Xi_{c,d}(\tau)$ in accordance with the laws (24) governing $\Theta^4{}_{0,0}(\tau)$. We shall prove that if $\Xi_{1,1}(\tau)\to 0$ as $\operatorname{Im}\tau$, $-\operatorname{Im}\tau'\to\infty$, then using the new coordinates,

$$\begin{cases} \tau = -\epsilon' u + v \\ \tau' = -\epsilon u + v, \end{cases} \tag{60}$$

the function $\Xi_{1,1}(\tau)$ *vanishes with order of magnitude* v^4 on $v=0$, (which is manifold (53) of course).

The proof is similar to Götzky's for $D = 5$ but we must outline the steps for $D = 8$ for later comparison (with $D = 12$). Call

$$\Gamma(u,v) = \Xi_{1,1}(\tau,\tau') \tag{61}$$

and call

$$\Gamma_t(u) = (\partial/\partial v)^t \Gamma(u,0). \tag{62a}$$

Then we define j as the integer ($j \geqq 0$) for which

$$\Gamma_0(u) \equiv \Gamma_1(u) = \cdots = \Gamma_j(u) \equiv 0; \qquad \Gamma_{j+1}(u) \not\equiv 0. \tag{62b}$$

Now, taking cognizance of the relations (deduced from (23)), we find

$$\begin{cases} \Xi_{1,1}(-\epsilon'^2/\tau) = N(\tau)^2 \Xi_{1,1}(\tau) \\ \Xi_{1,1}(\tau - \epsilon') = \Xi_{1,1}(\tau). \end{cases} \tag{63}$$

These relations can be differentiated with respect to v, (i.e., $\partial/\partial v = \partial/\partial\tau + \partial/\partial\tau'$) with the ensuing substitution $v = 0$. Then it is seen that for any t from 0 to j inclusive

$$\begin{cases} \Gamma_t(u+1) = \Gamma_t(u) \\ \Gamma_t(-1/u) = u^{2t+4}\Gamma_t(u) \\ \Gamma_t(u) \to 0 \qquad \text{as } \operatorname{Im} u \to \infty. \end{cases} \tag{64}$$

By accounting for singularities according to a well known method [2] we find $j \geqq 4$, since the smallest t for which relations (64) define a non-vanishing modular form occurs where $2t + 4 = 12$ and $\Gamma_4(u)$ is zero or else proportional to the twenty-fourth power of the Dedekind eta-function $\mathrm{H}(u)$, defined as

$$\mathrm{H}(u) = \exp(\pi i u/12) \prod_{n=1}^{\infty} (1 - \exp(2\pi i n u)). \tag{65}$$

The earlier $\Gamma_0(u) = \Gamma_1(u) = \Gamma_2(u) = \Gamma_3(u) \equiv 0$, of course.

Thus it follows, for $D = 8$, that for the $\Xi(\tau)$ just described $\Xi_{0,0}(\tau)/\Theta^4{}_{0,0}(\tau)$ generates a set of four modular functions under $\mathfrak{G}_D$ which *possess no finite singularities*. Hence by Theorem 4, this ratio is constant.

9. Results for $2^{\frac{1}{2}}$. There are several different types of results subsequently available for $D = 8$.

Basis Theorem. *The modular forms of dimension -2 belonging to the unimodular group $\mathfrak{G}_D$ form a vector space of dimension 1 with basis element $\psi_0(2\tau) = \Omega_3(\tau)$. Modular forms of dimension -2 belonging to the unimodular group $\mathfrak{H}_D$ form a vector space of dimension 2 (for example, with two basis elements, chosen from $\Theta^4{}_{0,0}(\tau)$, $\Omega_1(\tau)$, $\Omega_2(\tau)$, $\Omega_3(\tau)$).*

Proof. Clearly the first statement is a specialization of the second one. All we need note is that if $\Xi(\tau)$ is any modular form of dimension -2, then for some k, $\Xi_{1,1}(\tau)-\Omega_3(\tau)\cdot k$ vanishes, as $\operatorname{Im}\tau$, $-\operatorname{Im}\tau'\to\infty$, yielding a function which is a multiple of $\Theta^4{}_{0,0}(\tau)$ as shown in Section 8.

DECOMPOSITION THEOREM. *If $\mu=a+b2^{\frac{1}{2}}$ is a totally positive integer, then $A(\mu)$, the number of decompositions of μ into four squares, is given by the formulae* (2) *and* (3) *if b is even* (*and is zero if b is odd*).

Proof. By the method outlined above the linear relation is deduced from (51) and (52) using $C_D=1/48$:

$$\Theta^4{}_{0,0}(\tau)-4\psi_0(\tau)+4\psi_1(\tau)+4\psi_\infty(\tau)-24\psi_0(2\tau)=\Omega_{0,0}(\tau), \tag{66}$$

or more generally

$$\Theta^4{}_{d,e}(\tau)=4\psi_{1+(d,e)\eta}(\tau)+4\psi_{1/(d\eta)}(\tau)+4\psi_{e\eta}(\tau)-24\psi_0(2\tau). \tag{67}$$

The reader will note that the combination $\psi_0(\tau)+\psi_1(\tau)$ is the artifact that makes possible the exclusion of all $a+b2^{\frac{1}{2}}$ with b odd.

IDENTITIES ON THE ZERO MANIFOLD. If we define the functions $\Gamma_t(u)$ corresponding to $\Theta^4{}_{1,1}(\tau)=4\psi_1(\tau)+4\psi_\eta(\tau)+4\psi_\eta(2\tau)-24\psi_0(\tau)$, we find after considerable labor,

$$\begin{aligned}(2\pi i)^{-t}\Gamma_t(u)=2\sum_{a+b2^{\frac{1}{2}}\gg 0}&\exp[\pi iu(a-b)]b^t\{(-1)^a+(-1)^b\}D(a+b2^{\frac{1}{2}})\\ +8\sum_{a+b2^{\frac{1}{2}}\gg 0}&\exp[2\pi iu(a-b)](2b)^t(-)^aD(a+b2^{\frac{1}{2}})\\ -12\sum_{a+b2^{\frac{1}{2}}\gg 0}&\exp[2\pi iu(a-b)](2b)^tD(a+b2^{\frac{1}{2}}),\end{aligned} \tag{68}$$

where

$$D(\mu)=\sum_{(\nu)\mid\mu}|N(\nu)| \tag{69}$$

extended over all ideal divisors (ν) of μ. We interpret $0^0=1$ in (68) and obtain the more concise result, using $g=a-b>0$: For each $g>0$,

$$\begin{aligned}&\sum^{(b)}_{2g+b(1+2^{\frac{1}{2}})\gg 0}D(2g+b(1+2^{\frac{1}{2}}))(-1)^bb^t\\ &-\sum^{(b)}_{g+b(1+2^{\frac{1}{2}})\gg 0}D(g+b(1+2^{\frac{1}{2}}))\cdot(3-2(-1)^{g+b})(2b)^t=\begin{cases}0, & 0\leqq t\leqq 3\\ 48\tau(g), & t=4,\end{cases}\end{aligned} \tag{70}$$

where $\tau(g)$ is Ramanujan's notation for the coefficient of $\exp(2\pi igu)$ in $H^{24}(u)$.

The reader will note that Götzky has results of all three types, at least

implicitly, for $D=5$. To recognize the similarity of his identities we rewrite his result as

$$(71)\qquad \sum_{2g+b(1+5^{\frac12})/2\gg 0}^{(b)} D^*(2g+b(1+5^{\frac12})/2)(-1)^b b^t = \begin{cases} -24\tau(g), & t=4, \\ 0, & 0\leqq t\leqq 3 \end{cases}$$

where 0^0 is again interpreted as 1 and $D^*(\mu)$ is the sum function defined like D in (69), but limited to divisors (ν) for which $\mu/(\nu)$ is *odd*.

10. Zero manifold for $3^{\frac12}$ and Hecke's modular function. To deal with the case $D=12$ we substitute into formulae (22) and (17) the same value $\nu=a+b\eta$, where $\eta=1+3^{\frac12}$ this time (and not $2^{\frac12}$). Then using the new variables

$$(72)\qquad h_\eta=\exp[\pi i S(\eta\tau/12^{\frac12})] \qquad h_{\eta'}=\exp[\pi i S(\eta'\tau/12^{\frac12})],$$

we find, by symmetry

$$(73)\qquad \Theta_{1,c}=2\sum_{\substack{2a+2b+1>0\\2b+1>0}}(-1)^{ca}h_\eta^{[\frac12(a+2b+1)^2+(b+\frac12)^2]}h_{\eta'}^{[\frac12a^2+(b+\frac12)^2]}$$
$$+2\sum_{\substack{2a+2b+1>0\\2b+1>0}}(-1)^{c(a+1)}h_\eta^{[\frac12a^2+(b+\frac12)^2]}h_{\eta'}^{[\frac12(a+b+1)^2+(b+\frac12)^2]}.$$

Hence analogously with (59) we have

$$(74)\qquad \Theta_{1,c}/2(h_\eta^{\frac12}+(-)^c h_{\eta'}^{\frac12})=\sum_{\substack{2a+2b+1>0\\2b+1>0}}h_\eta^{[\frac12a^2+(b+\frac12)^2]}h_{\eta'}^{[\frac12a^2+(b+\frac12)^2]}f_{a,b},$$

where

$$(75)\qquad f_{a,b}=[h_\eta^{(2b+1)(2a+2b+1)/2}+(-)^c h_{\eta'}^{(2b+1)(2a+2b+1)/2}]/[h_\eta^{\frac12}+(-)^c h_{\eta'}^{\frac12}].$$

But $|h_\eta|\leqq 1$, $|h_{\eta'}|\leqq 1$ by equation (9). Hence using the argument of Section 7, we find that the functions $\Theta_{d,c}(\tau)$ for $D=12$ have as their only zeros in the fundamental domain the manifold of zeros of $\Theta_{1,1}(\tau)$,

$$(76)\qquad \tau+\tau'=0,$$

found by setting $h_\eta^{\frac12}=h_{\eta'}^{\frac12}$.

We next consider the new variables u, v defined by

$$(77)\qquad \begin{cases} \tau=u+v, \\ \tau'=-u+v, \end{cases}$$

and in particular we ask once more how $\Xi_{1,1}(\tau)$ will behave on the manifold $v=0$ if $\Xi(\tau)\equiv\Xi_{0,0}(\tau)$ is any modular form of dimension -2 belonging to

$\mathfrak{H}_D$ and such that $\Xi_{1,1}(\tau)$ vanishes at infinity. Accordingly, we set $\Gamma(u,v) = \Xi_{1,1}(\tau)$ and define

$$\Gamma_t(u) = (\partial/\partial v)^t \Gamma(u,0) \tag{78a}$$

so that as in (62b),

$$\Gamma_0(u) \equiv \Gamma_1(u) \equiv \cdots \equiv \Gamma_j(u) \equiv 0; \quad \Gamma_{j+1}(u) \not\equiv 0. \tag{78b}$$

From the relations

$$\begin{cases} \Xi_{1,1}(\tau+\epsilon) = \Xi_{1,1}(\tau) \\ \Xi_{1,1}(\tau+2) = \Xi_{1,1}(\tau) \\ \Xi_{1,1}(-1/\tau) = \Xi_{1,1}(\tau) N(\tau)^2, \end{cases} \tag{79}$$

we find as before $\Gamma_t(u)$ is a regular function with

$$\begin{cases} \Gamma_t(u+3^{\frac12}) = \Gamma_t(u) \\ \Gamma_t(-1/u) = \Gamma_t(u) u^{2t+4} \qquad (0 \leqq t \leqq j) \\ \Gamma_t(u) \to 0 \quad \text{as } \operatorname{Im} u \to \infty. \end{cases} \tag{80}$$

The relation (80) defines Hecke's modular function. The fundamental domain for the group generated by the operations indicated, namely $u \to u + 3^{\frac12}$, $u \to -1/u$, is given by the region

$$\mathfrak{R}_{3^{\frac12}} \begin{cases} -3^{\frac12}/2 \leqq \operatorname{Re} u \leqq +3^{\frac12}/2 \\ 1 \leqq |u|. \end{cases} \tag{81}$$

with proper identification of boundary points. The angles made at the points $u_0 = \pm 3^{\frac12}/2 + i/2$ are $30°$. Hence we can see, for example, that the region $\mathfrak{R}_{3^{\frac12}}$ is mapped on to the I-plane, with matched boundary points coinciding, by a function $I(u)$ with the following general description:

$$I(u) = \begin{cases} \text{const.} \exp(-2\pi i u/3^{\frac12}) + O(1) \ \text{ as } \operatorname{Im} u \to \infty \\ \text{const.}\ (u-i)^2 + 1 + O(u-i)^3 \ \text{ as } u \to i \\ \text{const.}\ (u-u_0)^6 + O(u-u_0)^7 \ \text{ as } u \to u_0. \end{cases} \tag{82}$$

Such a function is discussed in some detail in Raleigh's paper [14]. By the usual classical method [2] we can show that from the system (80),

$$\Gamma_t(u) = [I'(u)^{t+2}/(I(u)-1)^{t'} I(u)^{t''}] P_{t'''}(I(u)), \tag{83}$$

where

$$t' = [(t+2)/2],\ t'' = [5(t+2)/6], \text{ and } t''' = t' + t'' - t - 3.$$

Here t''' is the degree of the polynomial $P_{t'''}(I)$, with the usual convention that for $t''' < 0$, $P_{t'''}(I) \equiv 0$. We then discover that $j = 1$ or

$$(84)\qquad \begin{cases} \Gamma_0(u) \equiv \Gamma_1(u) \equiv 0 \\ \Gamma_2(u) = I'(u)^4/(I(u)-1)^2 I(u)^3 \cdot \text{const.}, \\ \Gamma_3(u) = I'(u)^5/I(u)-1)^2 I(u)^4 \cdot \text{const.}, \end{cases}$$

(whereas all the corresponding functions were necessarily zero for $D = 8$ and 5). Hence $\Xi_{1,1}(\tau)/\Theta_{1,1}(\tau)$ is not necessarily regular unless we can further stipulate that $\Gamma_2(u)$, $\Gamma_3(u)$ are $o(\exp(2\pi i u/3^{\frac{1}{2}}))$ as $\operatorname{Im} u \to \infty$.

11. Results for $3^{\frac{1}{2}}$. By following out the pattern of Section 9 as much as possible we obtain the following partial results for $D = 12$.

Basis Theorem. *The modular forms of dimension -2 belonging to the unimodular group $\mathfrak{G}_D$ form a vector space of dimension 1 or 2, with one basis element $\psi_0(2\tau) \equiv \Omega_3(\tau)$. Modular forms of dimension -2 belonging to the sub-group $\mathfrak{H}_D$ form a vector space of dimension 3 or 4 consisting of $\Theta^4_{0,0}(\tau)$ as well as two of the three functions $\Omega_1(\tau), \Omega_2(\tau), \Omega_3(\tau)$ (and possibly an additional function).*

Decomposition Theorem. *If $\mu = a + b3^{\frac{1}{2}}$ (b even) is a totally positive quadratic integer, then $A(\mu)$, the total number of decompositions of μ as the sum of four squares, is given by the "approximation"*

$$(85)\qquad A(\mu) = \tfrac{1}{2}B(\mu) + E(\mu),$$

where $B(\mu)$ is given by the same formulae (2), (3), but for $R(3^{\frac{1}{2}})$, and $E(\mu)$ is an "error" function, not identically zero. In fact, $E(1) = 4$. The modular form of dimension -2.

$$(86)\qquad Z_{0,0}(\tau) = \Theta^4_{0,0}(\tau) - \Omega_{0,0}(\tau) = \sum_{\mu \gg 0} E(\mu)\exp[\pi i S(\mu\tau/12^{\frac{1}{2}})]$$

is a "cusp form" in the sense that $Z_{d,c}(\tau) \to 0$ at ∞ for each d and c.

Both of the above theorems are related to each other. The factor $\frac{1}{2}$ in (85) comes from $C_{12} = 1/24$ (instead of $C_8 = 1/48$) in (51a). The coefficients $E(\mu)$ are majorized by $O(N(\mu)^q)$ for $q > 3/4$, hence easily by $o(B(\mu))$, according to a recent result of Gundlach [6]. Thus for $N(\mu)$ large enough, and possibly *all* $\mu = a + b3^{\frac{1}{2}} \gg 0$, ($b$ even), $A(\mu) > 0$, or four squares suffice. An "exact" study of $E(\mu)$, however seems more promising than economical majorizing constants.

From the fact that $E(\mu) \not\equiv 0$, it follows that the vector space of modular forms of dimension -2 belonging to the unimodular group $\mathfrak{H}_D$ is of dimension at least 3. The dimension can not exceed 4 since, according to the earlier section, the form $\Xi(\tau)$ completely determined by two projections $\Gamma_2(u)$, $\Gamma_3(u)$ if $\Xi_{1,1}(\tau)$ vanishes at ∞. It can, however, be ascertained from the first term that $Z_{0,0}(\tau) \not\equiv Z_{1,1}(\tau)$ hence $Z(\tau)$ is no part of the basis for $\mathfrak{G}_D$.

IDENTITY ON THE ZERO MANIFOLD. Expanding $\Omega_{1,1}(\tau)$, and forming $\Gamma_0(u) \equiv 0$ we find, for each $a \geqq 1$,

$$(87) \qquad \sum_{2a+b\,3^{\frac{1}{2}}>>0}^{(b)} D(2a+b\,3^{\frac{1}{2}})(-1)^b = \sum_{a+b\,3^{\frac{1}{2}}>>0}^{(b)} D(a+b\,3^{\frac{1}{2}})\,(3-2(-1)^{a+b}),$$

where $D(\mu)$ has, in regard to $R(3^{\frac{1}{2}})$, the meaning (69). The identity for $\Gamma_1(u) \equiv 0$ is completely trivial.

If we were to restrict ourselves to *symmetric* theta functions, as defined in Section 2 (above) we would find that the dimension of the vector space of modular forms of dimension -2 belonging to $\mathfrak{H}_D$ is exactly 3 and that of $\mathfrak{G}_D$ is exactly 2. It is possible to press for more intensive results (as did Maass [11] for $D=5$), involving modular forms of varying dimension when $D=8$ and $D=12$ but this will be the subject of a later paper. The usage of the theory of modular functions of the Klein and Hecke types seems to restrict us to these three values of D.

The author is indebted to the referee for calling attention to the valuable references [6], [8a], and [10] and for other suggestions.

UNIVERSITY OF ARIZONA.

REFERENCES.

[1] O. Blumenthal, "Über Modulfunktionen von mehreren Veränderlichen," *Mathematische Annalen*, vol. 56 (1903), p. 528.

[2] H. Cohn, "Some diophantine aspects of Modular Functions, I. Essential singularities," *American Journal of Mathematics*, vol. 71 (1949), pp. 403-416.

[3] ———, "Numerical study of the representation of a totally positive integer as the sum of quadratic integral squares," *Numerische Mathematik*, vol. 1 (1959), pp. 121-134.

[4] L. E. Dickson, *History of the theory of numbers*, vol. 3, New York, 1934.

[5] F. Götzky, "Über eine zahlentheoretische Anwendung von Modulfunktionen zweier Veränderlichen," *Mathematische Annalen*, vol. 100 (1928), pp. 411-437.

[6] K.-B. Gundlach, "Über die Darstellung der ganzen Spitzformen zu den Idealstufen der hilbertschen Modulgruppe und die Abschätzung ihrer Fourierkoeffizienten," *Acta Mathematica*, vol. 92 (1954), p. 314.

[7] E. Hecke, "Analytische Funktionen und algebraische Zahlen, II Teil," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 3 (1924), pp. 213-236.

[8] H. D. Kloosterman, "Thetareihen in total-reelen algebraischen Zahlkörpern," *Mathematische Annalen*, vol. 103 (1930), pp. 279-299.

[8a] M. Koecher, "Zur Theorie der Modulformen n-ten Grades, I," *Mathematische Zeitschrift*, vol. 59 (1954), pp. 399-416.

[9] F. J. Koksma, "Diophantische Approximationen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 4, part 4 (1936), p. 24.

[10] H. Maass, "Zur Theorie der automorphen Funktionen von n Veränderlichen," *Mathematische Annalen*, vol. 117 (1941), p. 567.

[11] ———, "Modulformen und quadratische Formen über dem quadratischen Zahlkörper $R(5^{\frac{1}{2}})$," *ibid.*, vol. 118 (1941), pp. 65-84.

[12] ———, "Über die Darstellung total positiver Zahlen des Körpers $R(5^{\frac{1}{2}})$ als Summe von drei Quadraten," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 14 (1941), pp. 185-191.

[13] H. Petersson, "Elliptische Modulfunktionen und automorphe Funktionen," *Naturforschung und Medizin in Deutschland*, 1939-1946, vol. 1, Reine Mathematik, pp. 243-276.

[14] J. Raleigh, "The Fourier coefficients of the invariants $j(2^{\frac{1}{2}}, \tau)$ and $j(3^{\frac{1}{2}}, \tau)$," *Transactions of the American Mathematical Society*, vol. 87 (1958), pp. 90-107.

[15] C. L. Siegel, "Über die analytische Theorie der quadratischen Formen," *Annals of Mathematics*, vol. 36 (1935), pp. 527-606; vol. 37 (1936), pp. 230-263; vol. 38 (1937), pp. 212-291.

COMPLEX ANALYTIC MAPPINGS OF RIEMANN SURFACES I.*

Dedicated to Professor Marston Morse

By Shiing-shen Chern.[1]

Introduction. The geometrical nature of the theory of functions in one complex variable is a well-known fact and has been particularly emphasized by L. Ahlfors (cf. Bibliography). It is also the most natural viewpoint, because complex function theory should be regarded as the first chapter of the theory of complex analytic mappings of complex manifolds and the classical study of value distributions is the study of the "size" of the image of a complex analytic mapping. We give in this paper a treatment, from a purely differential-geometric viewpoint, of complex analytic mappings of a Riemann surface (= one-dimensional complex analytic manifold) into a compact Riemann surface. In the case when the first Riemann surface is a compact one with a finite number of points deleted, we derive defect relations which generalize the classical relations of Nevanlinna-Ahlfors. In a subsequent paper we will consider the case when the first Riemann surface is a compact one with a finite number of points and a finite number of disks deleted. The paper is written for differential geometers, so that concepts currently in use in differential geometry are freely used and a minimum of function theory will be required. The explicit models of the Gaussian plane or the unit disk are avoided.

1. Hermitian metric on a Riemann surface. Let M be a Riemann surface. On M suppose there be an Hermitian metric, which is given, in terms of a local coordinate $z = x + iy$, by

$$ds^2 = h^2\, dz d\bar{z}, \tag{1}$$

where h is real and strictly positive. We suppose h to be of class C^∞ in the real local coordinates x, y. With the Hermitian metric it is possible to speak of the unit tangent vectors of M, the totality of which forms a circle bundle B over M. We denote by ψ the projection of B onto M.

* Received May 18, 1959.

[1] Work done under partial support of the National Science Foundation.

To the Hermitian metric there corresponds the associated two-form

$$\Omega = (i/2)h^2\, dz \wedge d\bar{z}. \tag{2}$$

It is a real-valued exterior two-form and is the element of area.

The Hermitian metric defines uniquely a connection in the bundle B, which can be described as follows: Relative to the local coordinate $z = x + iy$, the differentials dx, dy form a base in the cotangent space, and the vectors $\partial/\partial x$, $\partial/\partial y$ form its dual base in the tangent space. (Here $\partial/\partial x$ denotes the vector such that its directional derivative is the partial derivative with respect to x, and the same for $\partial/\partial y$.) As usual we introduce the complex vectors

$$\partial/dz = \tfrac{1}{2}(\partial/\partial x - i\partial/\partial y), \qquad \partial/\partial\bar{z} = \tfrac{1}{2}(\partial/\partial x + i\partial/\partial y). \tag{3}$$

Then the real unit vectors are given by

$$(e^{i\phi}\partial/\partial z + e^{-i\phi}\partial/\partial\bar{z})/h. \tag{4}$$

In $\psi^{-1}(U)$, where U is a neighborhood in which the local coordinate z is valid, z and ϕ can serve as local coordinates. The unit vector (4) defines one and only one complex-valued linear differential form

$$\omega_1 = e^{-i\phi}h\, dz, \tag{5}$$

characterized by the properties that it is of type $(1,0)$ and gives the value one when paired with (4). Let

$$\omega = -d\phi + i(d' - d'')\log h, \tag{6}$$

which is then a real-valued linear differential form in $\psi^{-1}(U)$. It can be verified that ω satisfies the equation

$$d\omega_1 = i\omega \wedge \omega_1. \tag{7}$$

and is the only real-valued linear differential form satisfying (7) and having the property that $\omega \equiv -d\phi$, $\mathrm{mod}\, dz, d\bar{z}$. This characterization of ω has the important implication that it is globally defined in B, independent of the choice of local coordinates. It is said to define a connection in B.

The exterior derivation of ω gives the important formula

$$d\omega = i\tfrac{1}{2}K\omega_1 \wedge \bar{\omega}_1 = K\Omega. \tag{8}$$

The coefficient K is a real-valued function in M, and is the curvature of the Hermitian metric. In terms of a local coordinate z one finds

$$K = -(4/h^2)\partial^2(\log h)/\partial z\partial\bar{z} \tag{9}$$

which is a well-known formula for the Gaussian curvature in isothermal parameters.

The Hermitian metric is called respectively Euclidean, hyperbolic, or elliptic, if the curvature K is constant and is $=0$ <0, or >0. It is well-known that a compact Riemann surface can always be given an Hermitian metric of constant curvature, the sign of the curvature being the same as the sign of its Euler characteristic.

2. Poisson's equation on a compact Riemann surface. Let Ω be a real-valued two-form of class C^1 on a compact Riemann surface M, such that $\int_M \Omega = c > 0$. An equation of the form

$$(1/\pi i)\, d'd''u = (1/c)\Omega \tag{10}$$

is called a Poisson equation. Since the equation can be written

$$(i/2\pi)\, d(d' - d'')u = (1/c)\Omega,$$

it does not have a smooth solution. The following theorem states that the equation has a solution with a logarithmic singularity at a given point of M:

THEOREM 1. *Let a be an arbitrary point on a campact Riemann surface M. Equation* (10) *has a solution $u(p,a)$, $p \in M$, having the following properties:* 1)*$u(p,a)$ is of class C^2 in $M - a$;* 2) *If z_a is a local coordinate at a such that $z_a = 0$ for a, then $u(p,a) - \log |z_a|$ is of class C^2 in a neighborhood of a.*

The function $u(p,a)$, whose existence is asserted in the theorem, is determined up to an additive constant, for the difference of two such functions is a harmonic function which is everywhere regular on M and is therefore a constant. We also remark that condition 2) is independent of the choice of the local coordinate z_a. For if z_a' is another such coordinate, we must have

$$z_a' = z_a f(z_a), \qquad f(0) \neq 0.$$

Then

$$\log |z_a'| = \log |z_a| + \log |f(z_a)|,$$

where $\log |f(z_a)|$ is regular at a.

To prove the theorem let $cv(p,a,b)$ be a harmonic function, which is regular in $M - a - b$, $a, b \in M$, $a \neq b$, and has the singularity $\log |z_a|$ at a and the singularity $-\log |z_b|$ at b, where z_b is a local coordinate at b such that $z_b = 0$ for b. This function $v(p,a,b)$ is defined up to an additive constant.

For $b=a$ we define $v(p,a,a)=\lim_{a\to b} v(p,a,b)=\text{const.}$ We shall prove that the function

$$u(p,a)=\int_M v(p,a,b)\Omega_b, \tag{11}$$

obtained by the integration of $v(p,a,b)$ with respect to $b\in M$, fulfills the conditions of our theorem.

In the first place we will show that the integral (11) converges. Suppose $p\neq a$. Let U_p be a neighborhood about p, with the local coordinate z_p and with $a\notin U_p$. We can write

$$\begin{aligned} u(p,a)=\int_{M-U_p} v(p,a,b)\Omega_b+\int_{U_p}\{v(p,a,b)+(1/c)\log|z_p-\zeta|\}\Omega_b \\ -(1/c)\int_{U_p}\log|z_p-\zeta|\,\Omega_b \end{aligned} \tag{12}$$

where ζ is the coordinate of b. The first two integrals in (12) are proper integrals, while the third integral is obviously convergent. Clearly the function $u(p,a)$ has the singularity $\log|z_a|$ at a.

To calculate $d'd''u(p,a)$ it suffices to restrict ourselves to the neighborhood U_p. The first two integrals in (12) can be differentiated under the integral sign and are annihilated by the operator $d'd''$, because the integrands are harmonic functions. In U_p let

$$\Omega_b=iF(\zeta)\,d\zeta\wedge d\bar{\zeta}.$$

Then we have

$$\begin{aligned}(1/c)\,d'd''\int_{U_p}-\log|z_p-\zeta|\,\Omega_b &= -(i/c)\,d'd''\int_{U_p}\log|z_p-\zeta|\,F(\zeta)\,d\zeta d\bar{\zeta} \\ &= -(i/c)\,dz_p\wedge d\bar{z}_p(\partial^2/\partial z_p\partial\bar{z}_p)\int_{U_p}\log|z_p-\zeta|\,F(\zeta)\,d\zeta d\bar{\zeta}.\end{aligned}$$

By a well-known computation (cf., for instance, I. G. Petrovsky, Lectures on Partial Differential Equations, p. 219) this is equal to $-(1/c)\pi F dz_p\wedge d\bar{z}_p=(\pi i/c)\Omega$. It follows that $d'd''u=(1/c)\pi i\Omega$, so that $u(p,a)$ is a solution of the Poisson equation (10). Similarly, it can be proved that $u(p,a)-\log|z_a|$ is regular at a and satisfies (10).

As an example we give the function $u(p,a)$ in the case that M is the Riemann sphere and Ω is the element of area. We consider the Riemann sphere as the complex projective line P_1, whose points have the homogeneous coordinates $Z=(z_0,z_1)$ and in which there is given an Hermitian scalar product

$$(Z,W)=\overline{(W,Z)}=z_0\bar{w}_0+z_1\bar{w}_1, \qquad W=(w_0,w_1). \tag{13}$$

We will write $|Z| = +(Z,Z)^{\frac{1}{2}} \geqq 0$. P_1 has an Hermitian metric of constant positive curvature 4, given by

$$ds^2 = [|Z|^2(dZ,dZ) - (Z,dZ)(dZ,Z)]/|Z|^4. \tag{14}$$

With this Hermitian metric, P_1 is also a metric space, and the distance $d(Z,W)$ between the points Z and W is given by the formula

$$\cos d(Z,W) = |(Z,W)|/|Z|\cdot|W|. \tag{15}$$

We now prove that the element of area is $\Omega = id'd'' \log |Z|$. In fact, let

$$Z_0 = Z/|Z|. \tag{16}$$

so that $(Z_0, Z_0) = 1$. Then

$$ds^2 = (dZ_0, dZ_0) = d\zeta_0 d\bar{\zeta}_0 + d\zeta_1 d\bar{\zeta}_1, \qquad Z_0 = (\zeta_0, \zeta_1), \tag{17}$$

and the element of area of P_1 relative to this Hermitian metric is

$$\Omega = \tfrac{1}{2}i(d\zeta_0 \wedge d\bar{\zeta}_0 + d\zeta_1 \wedge d\bar{\zeta}_1). \tag{18}$$

Let

$$\omega_{00} = (dZ_0, Z_0) = d\zeta_0\bar{\zeta}_0 + d\zeta_1\bar{\zeta}_1. \tag{19}$$

Then $\Omega = -\frac{1}{2}id\omega_{00}$. On the other hand, we have

$$2|Z|\,d|Z| = (dZ,Z) + (Z,dZ),$$

and

$$\begin{aligned}\omega_{00} = (dZ_0, Z_0) &= (1/|Z|)((1/|Z|)dZ - (Z/|Z|^2)d|Z|, Z)\\ &= (1/2|Z|^2)\{(dZ,Z) - (Z,dZ)\}\\ &= (d' - d'')\log|Z|.\end{aligned} \tag{20}$$

It follows that

$$\Omega = id'd'' \log |Z|. \tag{21}$$

An elementary computation gives $\int_{P_1} \Omega = \pi$, which is therefore the total area of P_1.

Let $A = (a_0, a_1)$, $A^\perp = (-\bar{a}_1, \bar{a}_0)$. Since $(Z, A^\perp)$ is holomorphic, $\log |(Z,A^\perp)|$ is zero under the operator $d'd''$, and we have

$$d'd'' \log(|(Z,A^\perp)|/|Z|\cdot|A|) = -d'd'' \log|Z| = i\Omega.$$

It follows that the function

$$u = \log(|(Z,A^\perp)|/|Z|\cdot|A|). \tag{22}$$

is a solution of the Poisson equation (10) and has a singularity $\log|z_A|$ at A, where z_A is a local coordinate at A. This is therefore the explicit expression of the function $u(p,a)$ whose existence was asserted by Theorem 1.

Since

$$d(Z,A)+d(Z,A^{\perp})=d(A,A^{\perp})=\pi/2,$$

we can write

$$u=\log\cos d(Z,A^{\perp})=\log\sin d(Z,A). \tag{23}$$

The quantity $\sin d(Z,A)$ is the length of the chord joining Z and A, when P_1 is realized as a sphere of diameter 1 in Euclidean space. This choice was made in the literature by Ahlfors and our discussion gives an intrinsic justification of the choice.

3. The first main theorem. Let D be a compact differentiable oriented domain bounded by a sectionally smooth curve C, and let $f\colon D\to M$ be a differentiable mapping. Suppose $f(\zeta_0)=a$, $\zeta_0\in D-C$, $a\in M$, and suppose that ζ_0 is the only point in a neighborhood of ζ_0 which is mapped into a by f. Then we can define an integer $n(\zeta_0,a)$, to be called the *order of f at* (ζ_0,a), which in a sense measures the number of times a neighborhood of a is covered by a neighborhood of ζ_0 under f. If a is such that $f^{-1}(a)$ is a finite set of points belonging to $D-C$, de define

$$n(a)=\sum_{\zeta\in f^{-1}(a)} n(\zeta,a). \tag{24}$$

The first main theorem expresses the difference of $n(a)$ and the area of $f(D)$ as an integral over the boundary curve C.

The definition of $n(\zeta_0,a)$ is as follows: Since f is a continuous mapping, there are coordinate neighborhoods U, V of ζ_0, a respectively, such that $f(U)\subset V$. Let $\zeta=\rho e^{i\theta}$, $z=re^{i\phi}$ be the local coordinates in U,V respectively, with ζ_0 and a having the coordinates $\zeta=0$, $z=0$. Denote by S_η and Σ_ϵ the circles $\rho=\eta=\text{const.}$ and $r=\epsilon=\text{const.}$ respectively, where η and ϵ are sufficiently small. There is a mapping g (called a retraction in topology): $V-a\to\Sigma_\epsilon$, which maps a point with the coordinate $z=re^{i\phi}$ into the point with the coordinate $\epsilon e^{i\phi}$. Since both D and M are oriented, the circles S_η, Σ_ϵ have induced orientations, and we use the same symbols to denote their fundamental cycles. Then $gf(S_\eta)$ is a cycle on Σ_ϵ and is homologous to an integral multiple of Σ_ϵ. This multiple, which can be shown to be independent of the various choices we have made, is defined to be the order $n(\zeta_0,a)$.

The order $n(\zeta_0, a)$ can be expressed by an integral formula. In fact, let $u(p,a)$ be a solution of (10) given by Theorem 1. Let

$$\lambda = (i/2\pi)(d' - d'')u, \tag{25}$$

so that $d\lambda = (1/c)\Omega$. Denote by $j\colon \Sigma_\epsilon \to V - a$ the identity mapping. Then

$$j^*\lambda = -\tfrac{1}{2\pi} d\phi + o(\epsilon),$$

where $o(\epsilon)$ denotes a differential form which tends to zero as $\epsilon \to 0$. It follows that

$$n(\zeta_0, a) = -\lim_{\epsilon\to 0} \int_{gf(S\eta)} j^*\lambda = -\lim_{\epsilon\to 0} \int_{jgf(S\eta)} \lambda.$$

We choose Ω in Theorem 1 to be the element of area of M. The first main theorem then follows by the application of Stokes' Theorem to the integral of $d\lambda$ over D. We state it as follows:

THEOREM 2. *Let D be a compact differentiable, oriented domain bounded by a sectionally smooth curve C, and let f be a differentiable mapping of D into a compact Riemann surface M. If $a \in M$ is such that $f^{-1}(a) \cap C = \emptyset$ and that $f^{-1}(a)$ is a finite set of points, then*

$$n(a) + \int_{f(C)} \lambda = (1/c) \int_{f(D)} \Omega = (1/c) v(D), \tag{26}$$

where $v(D)$ is the area of $f(D)$. In particular, if D has no boundary, then

$$n(a) = (1/c) v(D). \tag{27}$$

Remark. If f is orientation-preserving, then $v(D) \geqq 0$ and we have the following corollary of Theorem 2:

COROLLARY. *Let D be a compact oriented differentiable manifold of two real dimensions. Let $f\colon D \to M$ be an orientation-preserving differentiable mapping of D into a compact Riemann surface M. The image of D under f covers M completely, provided that the Jacobian of f does not vanish identically.*

4. The Gauss-Bonnet formula and the second main theorem. We now apply Stokes' Theorem to the formula (8). Suppose Δ be a domain in M bounded by a sectionally smooth curve γ. Since (8) is valid in the bundle B, and not in M, we have to "lift" Δ to B in order that Stokes' Theorem be applied. This is done by defining a differentiable field of unit vectors over Δ, such that it has a finite number of singularities in the interior of Δ and that the vectors at the points of γ point to the interior of Δ. According to a

known theorem in topology, the sum of indices at the singularities of such a vector field is equal to the Euler characteristic $\chi(\Delta)$ of Δ. From the local expression (6) of ω, the integral of ω along the vectors over a small circle about a singularity has as limit -2π times the index of the singularity, as the radius of the circle tends to zero. It follows by Stokes' Theorem that

$$(28) \qquad 2\pi\chi(\Delta) + \int_\gamma \omega = \int\int_\Delta K\Omega,$$

where the simple integral is over the vector field along γ. There is arbitrariness in the choice of the unit vector field along γ. If γ is smooth, it is natural to choose at each point $p \in \gamma$ the unit interior normal vector to γ at p. Then ω is equal to the element of arc of γ multiplied by its geodesic curvature. If γ is only sectionally smooth, then, corresponding to each corner, we have to add the exterior angle at that corner. Formula (28) is of course the well-known Gauss-Bonnet formula.

Consider now a compact complex analytic domain D bounded by a sectionally smooth curve C and a complex analytic mapping $f: D \to M$. If ζ is a local coordinate about a point $\zeta_0 \in D$, and z a local coordinate about the image point $a = f(\zeta_0)$, ζ_0 is called a stationary point or branch point, if $(dz/d\zeta)_{\zeta_0} = 0$. This condition is obviously independent of the choice of the local coordinates. If f is not a constant mapping, the stationary points are isolated. The Hermitian metric (1) in M induces the Hermitian metric $f^*(ds^2)$ in D, except at the stationary points. We will generalize the Gauss-Bonnet formula (28) to the domain D by taking account of the contributions arising from the isolated stationary points.

Suppose that the local coordinates ζ, z about ζ_0, a respectively are such that they vanish for ζ_0, a. The coordinate z of the image point $f(\zeta)$ is then given by a power series in ζ:

$$(29) \qquad z = a_m\zeta^m + a_{m+1}\zeta^{m+1} + \cdots, \qquad a_m \neq 0, m \geqq 1,$$

which is convergent in a neighborhood of $\zeta = 0$. The integer $m - 1$ is called the stationary index; a point of stationary index 0 is a regular point. About ζ_0 draw a small circle S_η of radius η and attach at each of its points the unit outward normal vector, so that it will be an inward normal vector of the complement domain. By (6) the integral of ω over this vector field tends to $-2\pi m$, as $\eta \to 0$.

If there are s stationary points, the Euler characteristic of the domain

D with the stationary points deleted is $\chi(D) - s$. Hence, if no stationary point lies on C, we have the formula

$$2\pi(\chi(D) - s) + \int_C \omega + 2\pi \sum_{1 \leqq i \leqq s} m_i = \int\int_{f(D)} K\Omega,$$

or

$$2\pi\chi(D) + \int_C \omega + 2\pi \sum_{1 \leqq i \leqq s} (m_i - 1) = \int\int_{f(D)} K\Omega, \tag{30}$$

where $m_i - 1$, $1 \leqq i \leqq s$, are the stationary indices.

If D has no boundary and is an n-leaved covering surface of M, then $\int\int_{f(D)} K\Omega = 2\pi n\chi(M)$, and (30) becomes

$$\chi(D) + \sum_{1 \leqq i \leqq s} (m_i - 1) = n\chi(M). \tag{31}$$

This is a well-known formula of A. Hurwitz.

5. Point at infinity. Let V_1 be a compact Riemann surface, and V the surface obtained from it by removing a finite number of points $v_1, \cdots, v_m$. Such a point v_k, $1 \leqq k \leqq m$, is called a point of infinity of V. Let $\zeta_k = r_k e^{i\theta_k}$ be a local coordinate at v_k, with $\zeta_k(v_k) = 0$, which will be fixed from now on. Denote by D_ϵ the domain obtained from V_1 by removing the disks $r_k \leqq \epsilon$, so that the boundary of D_ϵ consists of the m circles γ_k defined by $r_k = \epsilon$, and that D_ϵ tends to V as $\epsilon \to 0$. Consider a complex analytic mapping $f\colon V \to M$, where M is a compact Riemann surface, such that f does not send all points of V into one point. Then to every $a \in M$, the set $f^{-1}(a) \cap D_\epsilon$ is a finite set. If in addition $f^{-1}(a) \cap \gamma_k = \emptyset$, Theorem 2 gives

$$n(a, \epsilon) - \sum_{1 \leqq k \leqq m} \int_{\gamma_k} f^*\lambda = (1/c)v(\epsilon), \tag{32}$$

where c is the total area of M, $v(\epsilon)$ is the area of $f(D_\epsilon)$, and $n(a, \epsilon)$ is the number of times that a is covered by $f(D_\epsilon)$. The negative sign before the sum is caused by the fact that γ_k is to be oriented by increasing θ_k.

The integral in (32) can be simplified as follows: f being a complex analytic mapping, the dual mapping f^* commutes with both d' and d'', so that we have

$$f^*\lambda = (i/2\pi)(d' - d'')u(f(\zeta_k)).$$

Dropping the subscript k for the moment, consider the local coordinate $\zeta = re^{i\theta}$. If h is a real-valued differentiable function in r, θ, we have

$$dh = h_r\,dr + h_\theta\,d\theta = h_\zeta\,d\zeta + h_{\bar\zeta}\,d\bar\zeta$$
$$= (h_\zeta e^{i\theta} + h_{\bar\zeta} e^{-i\theta})\,dr + ir\,(h_\zeta e^{i\theta} - h_{\bar\zeta} e^{-i\theta})\,d\theta,$$

so that

$$h_r = h_\zeta e^{i\theta} + h_{\bar\zeta} e^{-i\theta}$$

and

$$(d' - d'')h = h_\zeta\,d\zeta - h_{\bar\zeta}\,d\bar\zeta = (\cdot\cdot\cdot)\,dr + irh_r\,d\theta. \tag{33}$$

It follows from (32) that

$$n(a,\epsilon) + \tfrac{1}{2\pi}\epsilon \sum_k \int_0^{2\pi} (\partial u/\partial\epsilon)\,d\theta_k = (1/c)\,v(\epsilon). \tag{34}$$

This leads us to put

$$T(\epsilon) = (1/c) \int_\epsilon^{\epsilon_0} v(t)\,dt/t, \qquad N(a,\epsilon) = \int_\epsilon^{\epsilon_0} n(a,t)\,dt/t, \tag{35}$$

where ϵ_0 is a small constant, with $\epsilon < \epsilon_0$. A standard argument will justify the relation

$$\int_0^{2\pi} (\partial u/\partial\epsilon)\,d\theta_k = (\partial/\partial\epsilon) \int_0^{2\pi} u\,d\theta_k,$$

and we get, by integrating the relation (34),

$$N(a,\epsilon) + \tfrac{1}{2\pi} \sum_k \int_0^{2\pi} u(\epsilon e^{i\theta_k}, a)\,d\theta_k \Big|_\epsilon^{\epsilon_0} = T(\epsilon),$$

or

$$N(a,\epsilon) - \tfrac{1}{2\pi} \sum_k \int_0^{2\pi} u(\epsilon e^{i\theta_k}, a)\,d\theta_k = T(\epsilon) + \text{const.} \tag{36}$$

Since $u(p,a)$ differs by a regular function from $\log|z_a|$ in a neighborhood of a and since M is compact, $u(p,a)$ has an upper bound, and we have the fundamental inequality

$$N(a,\epsilon) < T(\epsilon) + \text{const.} \tag{37}$$

This shows that the *order function* $T(\epsilon)$ dominates $N(a,\epsilon)$ for all $a \in M$.

We now derive an explicit formula for the second main theorem as applied to the domain D_ϵ. Let U_k be a coordinate neighborhood in V_1 in which the local coordinate ζ_k is valid. The restriction of f gives a mapping $f\colon U_k - v_k \to M$. Let $f^*\Omega = \sigma_k{}^2 r_k dr_k \wedge d\theta_k$, so that $\sigma_k{}^2$ is the ratio of the two elements of area. Since f is complex analytic, we have $\sigma_k{}^2 \geqq 0$ and $\sigma_k = 0$ if and only if the point is a stationary point. To the points of $U_k - v_k$ we

attach the radial vectors $\partial/\partial r_k$ and we will compute ω for this family of radial vectors. Since $\partial/\partial r_k = e^{i\theta_k}\partial/\partial\zeta_k + e^{-i\theta_k}\partial/\partial\bar{\zeta}_k$, we get, from (4) and (6),

$$\omega = -d\theta_k + i(d' - d'')\log\sigma_k. \tag{38}$$

If γ_k is free from stationary points, the restriction of ω to γ_k is

$$\omega = -(1 + \epsilon(\partial/\partial\epsilon)\log\sigma_k)\,d\theta_k.$$

Let $n_1(\epsilon)$ be the sum of stationary indices in D_ϵ. Then the second main theorem (30) reduces to the formula

$$\chi(V_1) - m + \tfrac{1}{2\pi}\sum_k \int_0^{2\pi}(1 + \epsilon(\partial/\partial\epsilon)\log\sigma_k)\,d\theta_k + n_1(\epsilon) = \tfrac{1}{2\pi}\iint_{f(D\epsilon)} K\Omega,$$

or

$$\chi(V_1) + (\epsilon/2\pi)\sum_k \int_0^{2\pi}(\partial/\partial\epsilon)\log\sigma_k\,d\theta_k + n_1(\epsilon) = \tfrac{1}{2\pi}\iint_{f(D\epsilon)} K\Omega. \tag{39}$$

In particular, if the Gaussian curvature K is constant, we have

$$\chi(V_1) + (\epsilon/2\pi)\sum_k \int_0^{2\pi}(\partial/\partial\epsilon)\log\sigma_k\,\mathrm{d}\theta_k + n_1(\epsilon) = (K/2\pi)v(\epsilon). \tag{40}$$

We put

$$N_1(\epsilon) = \int_\epsilon^{\epsilon_0} n_1(t)\,dt/t, \tag{41}$$

and remark that the relation

$$\int_0^{2\pi}(\partial/\partial\epsilon)\log\sigma_k\,d\theta_k = (\partial/\partial\epsilon)\int_0^{2\pi}\log\sigma_k\,d\theta_k$$

can be justified by a standard argument. Integration of the above equation then gives

$$\chi(V_1)(\log\epsilon_0 - \log\epsilon) + \tfrac{1}{2\pi}\sum_k \int_0^{2\pi}\log\sigma_k\,d\theta_k\Big|_\epsilon^{\epsilon_0} + N_1(\epsilon)$$
$$= (cK/2\pi)T(\epsilon) = \chi(M)T(\epsilon)$$

or

$$-\chi(V_1)\log\epsilon - \tfrac{1}{2\pi}\sum_k \int_0^{2\pi}\log\sigma_k\,d\theta_k + N_1(\epsilon) = \chi(M)T(\epsilon) + \text{const.} \tag{42}$$

This is the integrated form of the second main theorem.

6. Generalization of the defect relation of Nevanlinna-Ahlfors. Let $a_1, \cdots, a_q$ be a finite set of points in M, and let

$$\rho(p) = c_1\Big(\prod_{1\leq i\leq q}\exp u(p, a_i)\Big)^{-2\lambda}, \qquad 0 < \lambda < 1, \tag{43}$$

where c_1 is a constant. Then the integral $\int_M \rho(p)\Omega$, where Ω is the element of area of M, is convergent. For the points where the convergence of the integral should be checked are the points a_i and at a_i the integrand is $1/|z_{a_i}|^{2\lambda}$ times a bounded function. We choose the constant c_1 so that $\int_M \rho(p)\Omega = 1$.

We now integrate the inequality (37) over M with respect to the density $\rho(p)\Omega$. Noticing that $n(p,t)$ is the number of times that the point p is covered by $f(D_t)$, we get

$$\int_\epsilon^{\epsilon_0} (dt/t) \int_{D_t} \rho f^*\Omega < T(\epsilon) + \text{const.}$$

Now

$$\int_{D_t} \rho f^*\Omega > \int_{D_t - D_{\epsilon_0}} \rho f^*\Omega - \int_t^{\epsilon_0} r\, dr \sum_k \int_0^{2\pi} \rho\sigma_k{}^2\, d\theta_k$$

so that we have

$$\int_\epsilon^{\epsilon_0} (dt/t) \int_t^{\epsilon_0} r\, dr \sum_k \int_0^{2\pi} \rho\sigma_k{}^2\, d\theta_k < T(\epsilon) + \text{const.} \tag{44}$$

LEMMA. *The inequality*

$$\int_\epsilon^{\epsilon_0} (dt/t) \int_t^{\epsilon_0} e^{\phi(r)} r\, dr < S(\epsilon) \tag{45}$$

implies

$$\phi(\epsilon) \leqq \kappa^2 \log S(\epsilon) - 2\log \epsilon, \qquad \kappa > 1, \tag{46}$$

except for a set of intervals in $(0, \epsilon_0)$ *for which* $\int_\epsilon^{\epsilon_0} d\log\epsilon < \infty$.

Proof. Let $A(\epsilon)$ be a decreasing positive-valued C^1-function and $\Lambda(\epsilon)$ be a positive-valued continuous function, both defined in the interval $0 < \epsilon \leqq \epsilon_0$. Suppose

$$-A'(\epsilon) > A(\epsilon)^\kappa \Lambda(\epsilon), \qquad \kappa > 1.$$

Then

$$\Lambda(\epsilon) < -(1/A^\kappa)\, dA/d\epsilon = (1/(\kappa-1))\, dA^{1-\kappa}/d\epsilon$$

and

$$\int_\epsilon^{\epsilon_0} \Lambda(\epsilon)\, d\epsilon < (1/(\kappa-1))(A(\epsilon_0)^{1-\kappa} - A(\epsilon)^{1-\kappa}),$$

so that the integral at the left-hand side converges as $\epsilon \to 0$. Hence, with the exception of a set of $0 < \epsilon \leqq \epsilon_0$ for which the integral $\int_0^{\epsilon_0} \Lambda(\epsilon)\, d\epsilon$ converges, we have

$$-A'(\epsilon) \leqq A(\epsilon)^\kappa \Lambda(\epsilon),$$

whence

$$\log(-A') \leqq \kappa \log A(\epsilon) + \log \Lambda(\epsilon).$$

It follows that

$$\phi(\epsilon) + \log \epsilon \leqq \kappa \log \int_{\epsilon}^{\epsilon_0} e^{\phi(r)} r\, dr + \log \Lambda(\epsilon),$$

$$-\log \epsilon + \log \int_{\epsilon}^{\epsilon_0} e^{\phi(r)} r\, dr \leqq \kappa \log \int_{\epsilon}^{\epsilon_0} (dt/t) \int_{t}^{\epsilon_0} e^{\phi(r)} r\, dr + \log \Lambda(\epsilon),$$

so that

$$\phi(\epsilon) + (1-\kappa)\log \epsilon \leqq \kappa^2 \log S(\epsilon) + (1+\kappa)\log \Lambda(\epsilon).$$

Choose $\Lambda(\epsilon) = 1/\epsilon$; then

$$\phi(\epsilon) \leqq \kappa^2 \log S(\epsilon) - 2 \log \epsilon,$$

with the exception of a subset of $0 < \epsilon \leqq \epsilon_0$ for which $\int_0^{\epsilon_0} d \log \epsilon < \infty$. This completes the proof of the lemma.

Following H. Weyl we use the notation $\|$ to denote that an inequality is not universally valid, but only under the exception stated in the lemma.

It follow from the lemma that

$$(47)\,\| \qquad \log \sum_k \int_0^{2\pi} \rho \sigma_k^2\, d\theta_k \leqq \kappa^2 \log(T(\epsilon) + C) - 2 \log \epsilon,$$

where C is a constant. By the concavity of the logarithm we have

$$(48)\,\| \qquad \begin{aligned} \tfrac{1}{2\pi} \sum_k \int_0^{2\pi} \log(\rho\sigma_k^2)\, d\theta_k &\leqq \sum_k \log\{ \tfrac{1}{2\pi} \int_0^{2\pi} \rho \sigma_k^2\, d\theta_k \} \\ &\leqq m \log\{ (1/2m\pi) \sum_k \int_0^{2\pi} \rho \sigma_k^2\, d\theta_k \} \\ &\leqq m\kappa^2 \log(T(\epsilon) + C) - 2\, m \log \epsilon + O(1). \end{aligned}$$

On the other hand, by means of (36), (42), (43), the left-hand side of this inequality can be transformed as follows:

$$\begin{aligned} &\tfrac{1}{2\pi} \sum_k \int_0^{2\pi} \log(\rho\sigma_k^2)\, d\theta_k \\ &\quad = \tfrac{1}{2\pi} \sum_k \int_0^{2\pi} \log \rho\, d\theta_k + (2/2\pi) \sum_k \int_0^{2\pi} \log \sigma_k\, d\theta_k \\ &\quad = \text{const} - (2\lambda/2\pi) \sum_k \sum_{1 \leqq j \leqq q} \int_0^{2\pi} u(\epsilon e^{i\theta_k}, a_j)\, d\theta_k \\ &\qquad\qquad - 2\chi(V_1)\log \epsilon + 2N_1(\epsilon) - 2\chi(M)T(\epsilon) \\ &\quad = \text{const} + 2(\lambda q - \chi(M))T(\epsilon) - 2\lambda \sum_{1 \leqq j \leqq q} N(a_j, \epsilon) \\ &\qquad\qquad - 2\chi(V_1)\log \epsilon + 2N_1(\epsilon). \end{aligned}$$

Letting $\lambda \to 1$ and using (48), we get

$$(49)\| \quad \begin{aligned}(q-\chi(M))T(\epsilon) - \sum_{1 \leq j \leq q} N(a_j, \epsilon) + N_1(\epsilon) \\ \leq \tfrac{1}{2} m\kappa^2 \log(T(\epsilon) + C) + \chi(V)\log\epsilon + O(1).\end{aligned}$$

For $a \in M$ we define the defect of the mapping f by

$$(50) \qquad \delta(a) = 1 - \overline{\lim}(N(a,\epsilon)/T(\epsilon)).$$

Then $0 \leq \delta(a) \leq 1$ and $\delta(a) = 1$, if the point a does not belong to the image $f(V)$.

THEOREM 3. *Let V be a non-compact Riemann surface which is obtained from a compact one by the deletion of a finite number of points. Let M be a compact Riemann surface and let $f\colon V \to M$ be a non-constant complex analytic mapping. Let $a_1, \cdots, a_q \in M$. If $\chi(V) \geq 0$ or if $\underline{\lim}(-\log\epsilon/T(\epsilon)) = 0$, then*

$$(51) \qquad \sum_{1 \leq j \leq q} \delta(a_j) \leq \chi(M).$$

Condition (51) means that $\sum \delta(a_j) \leq 2$, when M is the Riemann sphere. When M is a complex torus, then (51) implies that all the defects are zero. If M is of genus > 1, (51) is a contradiction, so that such a mapping f does not exist.

The theorem follows immediately from (49), for the latter implies, as $N_1(\epsilon) \geq 0$,

$$\sum_j \delta(a_j) \leq \chi(M) - \chi(V)\underline{\lim}(-\log\epsilon/T(\epsilon)).$$

To clarify Theorem 3, we have to study the meaning of the condition

$$\underline{\lim}(-\log\epsilon/T(\epsilon)) = 0.$$

Suppose $\underline{\lim}(-\log\epsilon/T(\epsilon)) \neq 0$. Then $\overline{\lim}(T(\epsilon)/-\log\epsilon) = \mathrm{b} < \infty$, and we have $T(\epsilon) = O(-\log\epsilon)$.

THEOREM 4. *Suppose V and M be Riemann surfaces such that M is compact and V is obtained from a compact one V_1 by the deletion of a finite number of points. Let $f\colon V \to M$ be a non-constant complex analytic mapping such that $T(\epsilon) = O(-\log\epsilon)$. Then f can be extended into a complex analytic mapping of V_1 into M.*

Let $v \in V_1$ be a point at infinity of V. It follows from (37) and the condition $T(\epsilon) = O(-\log\epsilon)$ that $n(a,\epsilon)$, $a \in M$, is bounded in a neighborhood of v. Consider first the case that $M = M_0$ is the Riemann sphere. The

mapping $f\colon V \to M_0$ defines a meromorphic function in a neighborhood of v with an isolated singularity at v. Since $n(a, \epsilon)$ is bounded, the image of $U - v$ under f will omit a finite number of points of M_0, provided that the neighborhood U is small enough. It follows that v is a removable singularity or a pole, i.e., that the mapping f can be extended to a complex analytic mapping of V_1 into M_0.

In the general case that M is an arbitrary compact Riemann surface let M_0 be the Riemann sphere and let $g\colon M \to M_0$ be a covering of M_0, possibly ramified. The composition $gf\colon V \to M_0$ is a complex analytic mapping for which the condition $T(\epsilon) = O(-\log\epsilon)$ is still fulfilled. Hence gf can be extended to a complex analytic mapping of V_1 into M_0. It follows that, if U is small enough, $f(U - v)$ will belong to a finite number of neighborhoods of M, which could moreover be arbitrarily small. Since this is true for any covering g, this is possible only when $f(U - v)$ belongs to one coordinate neighborhood of M. In terms of local coordinates f defines a bounded holomorphic function. Thus the singularity v is removable, and our proof of Theorem 4 is complete.

It seems reasonable to call a non-constant complex analytic mapping $f\colon V \to M$ transcendental, if $T(\epsilon) \neq O(-\log\epsilon)$. Then we can say that a transcendental mapping f satisfies the defect relation (51).

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] L. V. Ahlfors, "Geometrie der Riemannschen Flächen," *Comptes Rendus du Congrès International des Mathematiciens*, Oslo 1936, pp. 239-248.

[2] ———, "Über die Anwendung differentialgeometrischer Methoden zur Untersuchung von Überlagerungsflächen," *Acta Societatis Scientiarum Fennicae*, Nova Series A, Tom II, No. 6, pp. 1-17 (1937).

[3] G. af Hällström, "Über meromorphe Funktionen mit mehrfach zusammenhängenden Existenzgebieten," *Acta Academiae Aboensis, Mathematica et Physica*, XII, vol. 8 (1939), pp. 1-100.

A PROOF OF A THEOREM ON MODULAR FUNCTIONS.*

By Hans Rademacher.

The following theorem is fundamental in the theory of modular functions:

Theorem. *A modular function $\psi(\tau)$ belonging to a modular congruence subgroup modulo N and which is regular and bounded in the half-plane $\Re(\tau)>0$ is a constant.*

This theorem is usually proved by a contour integration around the fundamental region of $\psi(\tau)$ and requires a study of the structure of this region, in particular of its vertices and cusps. The following proof avoids all this. All we need is Jensen's inequality:

Lemma. *If $f(z)$ is regular in the interior of the unit circle, $f(0)\neq 0$ and $|f(z)|<M$, then for any zeros $z_1, z_2, \cdots, z_n$ of $f(z)$ we have*

$$|z_1 z_2 \cdots z_n| \geqq |f(0)|/M.$$

This lemma implies the simple

Corollary. *If $g(z)$ is regular and bounded in $|z|<1$, and if z_ν, $\nu=1,2,3,\cdots, z_\nu\neq 0$, are zeros of $g(z)$ in the unit circle so that*

$$\prod_{\nu=1}^{\infty} |z_\nu| = 0,$$

then $g(z)$ vanishes identically.

For the proof of the theorem it suffices to consider the principal subgroup $\Gamma(N)$ modulo N

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{N},$$

which is contained as subgroup in any congruence subgroup modulo N.

Now $\Gamma(N)$ has $\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}$ as an element so that we can restrict our considerations to the domain

$$(1) \qquad -\tfrac{1}{2}N \leqq \Re(\tau) < \tfrac{1}{2}N, \qquad \Re(\tau) > 0.$$

* Received June 18, 1959.

Since $\psi(\tau+N)=\psi(\tau)$ we can introduce

$$(2) \qquad \tau=(N/2\pi i)\log z$$

and obtain a uniquely defined function $g(z)=\psi(\tau)$ for $|z|<1$, the unit circle being the image of (1) by means of (2).

The function $g(z)$ is *bounded* and *regular* in $|z|<1$. The possible singularity at $z=0$, stemming from $\tau=i\infty$, is a removable one since $g(z)$ is bounded.

Let us consider now the images of $\tau=i$ under the group $\Gamma(N)$. For any pair of natural numbers $c\equiv 0$, $d\equiv 1 \pmod N$ and $(c,d)=1$ there exists exactly one pair a,b so that

$$M=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(N)$$

and that

$$M(i)=(ai+b)/(ci+d)=(i+(ac+bd))/(c^2+d^2)=\tau_{c,d}$$

has moreover the property

$$(3) \qquad -\tfrac{1}{2}N \leqq \Re M(i)=(ac+bd)/(c^2+d^2)<\tfrac{1}{2}N.$$

Indeed, we have to solve the diophantine equation

$$(4) \qquad ad-bc=1$$

for a and b with the conditions

$$a\equiv 1, \qquad b\equiv 0 \pmod N.$$

Now (4) implies

$$ad\equiv 1 \pmod{Nc},$$

whch is solvable since $(d,Nc)=1$. Let a_0 be a particular solution. All solutions are then of the form

$$a=a_0+mNc,$$

and the remaining b in (4) is given by

$$b=(a_0d-1)/c+mNd.$$

A short computation shows that (3) is expressed as

$$-\tfrac{1}{2}N\leqq a_0/c-d/c(c^2+d^2)+mN<\tfrac{1}{2}N,$$

which is satisfied by exactly one integer m, so that a and b and thus $\tau_{c,d}$ are now uniquely determined.

We have

$$\Im M(i)=\Im\tau_{c,d}=1/(c^2+d^2).$$

In the z-plane these values correspond to

$$|z_{c,d}| = \exp(-2\pi/N(c^2+d^2)).$$

We want to prove

$$\prod |z_{c,d}| = 0$$

where the multiplication is extended over all $c, d > 0$ with $c \equiv 0$, $d \equiv 1$ $(\operatorname{mod} N)$, $(c, d) = 1$. Translated back into τ-coordinates our task is to prove that

$$\sum 1/(c^2+d^2) \tag{5}$$

is divergent.

But this is simple. Let us consider not the full series (5) but to any $d \equiv 1 \pmod N$ only those $c = \gamma N$ with $0 < \gamma \leqq d$ and, of course,

$$(c, d) = (\gamma, d) = 1.$$

These are $\phi(d)$ in number. For such a pair c, d we have $c^2 + d^2 \leqq (N^2+1)d^2$, so that the series (5) is estimated from below by

$$\sum 1/(c^2+d^2) > (N^2+1)^{-1} \sum_{\substack{d>0 \\ d\equiv 1 (\operatorname{mod} N)}} \phi(d)/d^2.$$

But the latter series is indeed divergent. This can be seen in several ways, for instance by means of the trivial estimate

$$\phi(n)/n \geqq \log 2/\log 2n$$

(obtained in the following way:

$$\phi(n)/n = \prod_{p|n} (1-1/p) \geqq (1-\tfrac{1}{2})(1-\tfrac{1}{3}) \cdots (1-1/(r+1)) = 1/(r+1)$$

with

$$n \geqq \prod_{p|n} p = p_1 p_2 \cdots p_r \geqq 2^r,$$

$$r \leqq \log n/\log 2, \qquad r+1 \leqq \log 2n/\log 2).$$

And the series $\sum_{d>0} 1/d \log 2d$, $d \equiv 1 \pmod N$ is divergent.

This shows that the spots $z_{c,d}$ where $g(z) - \psi(i)$ vanishes make $\prod(z_{c,d})$ diverge to zero. Then, in view of the Corollary, $g(z)$ is identically equal to $\psi(i)$, and $\psi(\tau)$ is constant. Q. E. D.

THE UNIVERSITY OF PENNSYLVANIA.

TAME COVERINGS AND FUNDAMENTAL GROUPS OF ALGEBRAIC VARIETIES.*

Part IV: Product Theorems.

By Shreeram Abhyankar.

Introduction. Let V be a nonsingular algebraic variety over an algebraically closed ground field k of characteristic p; let $W_1, \cdots, W_s, H_1, \cdots, H_t$ be distinct irreducible subvarieties of V of codimension one; let

$$W = W_1 \cup \cdots \cup W_s \text{ and } H = H_1 \cup \cdots \cup H_t.$$

In Section 3 it is proved that under certain conditions relating H and W there exists an exact sequence

$$0 \to \bar{\pi}'(V - W - H) \to \bar{\pi}'(V - W) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \to G \to 0,$$

where $\bar{\pi}'(V - W)$ (respectively: $\bar{\pi}'(V - W - H)$) is the galois group over K of the compositum of all finite extensions of $k(V)$ which are tamely ramified over V and for which the branch locus over V is contained in W (respective: in $W \cup H$), where $\bar{Z}^p$ occurs t times, where $\bar{Z}^p$ denotes the inverse limit of the inverse systems of all finite factor groups (of order prime to p in case $p \neq 0$) of an infinite cyclic group, and where G is a finite abelian group. As in [A3][1], the main technique is that of removing tame ramification through cyclic compositums [A3, Section 2]. Using the results of [A2], some corollaries of this result are deduced.

Let S and T be projective lines over k, let $S_1, \cdots, S_s$ $(s > 0)$ and $T_1, \cdots, T_t$ be distinct points on S and T respectively. Let

$$S^* = S - S_1 - \cdots - S_s \text{ and } T^* = T - T_1 - \cdots - T_t.$$

In Section 4 (Remark 4) it is shown that for $t = 1$ or 2 we have

$$\bar{\pi}^*(S^* \times T^*) \approx \bar{\pi}^*(S^*) \times \bar{\pi}^*(T^*)$$

* Received May 18, 1959.

[1] Numbers in square brackets refer to the References at the end of the paper. We take this opportunity to correct the following misprint: In the third line of Proposition 1 of [A3], the second K should be K^*.

and that such a product theorem does not hold for $\bar{\pi}'$. Some other results of a similar nature are derived (Theorem 4 and Corollary of Theorem 2). As a by-product of the results of Section 4, it is found (Remarks 1 and 3) that tame ramification is a birational non-invariant even for non-singular surfaces, and that $\bar{\pi}'$ is a birational non-invariant (whereas $\bar{\pi}^*$ and $\bar{\pi}$ are obviously invariant), and that one does not have a theorem of "purity of non-tame branch loci" at a simple point.

In Section 2 are given some lemmas to be used in this and succeeding papers of this series; and this includes a positive case of invariance of tame ramification for a certain type of birational transformations (Lemma 6).

1. Conventions, notations and definitions. Besides the conventions, notations and definitions introduced in [A1, 2, 3], we shall use the following ones:

(A). k will denote an algebraically closed ground field of characteristic p.

(B). Let a be a positive integer. By the expressions "a is prime to p," "let b be the part of a prime to a", $\cdots$, etc., we shall respectively mean, "a is prime to p in case $p \neq 0$ (vacuous statement in case $p = 0$)," "let $b = a$ in case $p = 0$ and let $b =$ the greatest factor of a prime to p in case $p \neq 0$", $\cdots$, etc.

(C). Let (R, M) be a regular local domain with quotient field K. For nonzero elements a_1 and a_2 of R, set $v(a_1/a_2) = n_1 - n_2$, where n_j is the largest integer such that $a_j \in M^{n_j}$; in other words, n_j is the R-leading degree $\lambda_R(a_j)$ of a_j. Then v is a real discrete valuation of K. We shall call v *the R-adic divisor of K*. If R is the quotient ring of a point P on an irreducible variety V, then we shall also call v *the P-adic divisor of V*.

(D). Let V be a normal projective irreducible algebraic variety over k. Let $W_1, \cdots, W_s, H_1, H_2$ be divisors on V. If H_1 is linearly equivalent to H_2 plus a sum of integral multiples of $W_1, \cdots, W_s$, then we shall say that H_1 is linearly equivalent to $H_2 \bmod W_1, \cdots, W_s$, and we shall express this by writing:

$$H_1 \equiv H_2 (\bmod W_1, \cdots, W_s).$$

(E). For the purposes of this and succeeding papers of the series it will be convenient to elaborate and extend the definitions of Section 10 of [A1] as follows: Let V be an n dimensional normal projective irreducible algebraic variety over k, let W be a subvariety of V, let $K = k(V)$, let Q be a

set of valuations of K/k of k-dimension $n-1$, and let Ω be an algebraic closure of K. We set:

$\Omega(V-W+Q)$ = the family of finite separable extensions L of K in Ω such that $\Delta(L/V)\subset W$ and such that no element of Q is ramified in L.

$\Omega'(V-W+Q)$ = the family of members of $\Omega(V-W+Q)$ which are tamely ramified over V.

$\Omega^*(V-W+Q)$ = the family of members of $\Omega(V-W+Q)$ such that the degree over K of the least galois extension of K containing L is prime to p.

$\Omega_g(V-W+Q)$ = the family of members of $\Omega(V-W+Q)$ which are galois over K.

$\Omega'_g(V-W+Q)$ = the family of members of $\Omega'(V-W+Q)$ which are galois over K.

$\Omega^*_g(V-W+Q)$ = the family of members of $\Omega_g(V-W+Q)$ whose degree over K is prime to p.

$\bar\Omega(V-W+Q), \bar\Omega'(V-W+Q), \bar\Omega^*(V-W+Q)$ = the compositum of all the members respectively of $\Omega(V-W+Q)$, $\Omega'(V-W+Q)$, $\Omega^*(V-W+Q)$.

Again, under the natural homomorphisms, the galois groups over K of the members respectively of (1) $\Omega_g(V-W+Q)$, (2) $\Omega'_g(V-W+Q)$, (3) $\Omega^*_g(V-W+Q)$ form group towers and we denote them respectively by (1) $\pi(V-W+Q)$, (2) $\pi'(V-W+Q)$, (3) $\pi^*(V-W+Q)$, and we call them respectively, (1) *the fundamental group tower of* $V-W+Q$, (2) *the tame fundamental group tower of* $V-W+Q$, (3) *the reduced fundamental group tower of* $V-W+Q$. Also the inverse limits respectively of (1) $\pi(V-W+Q)$, (2) $\pi'(V-W+Q)$, (3) $\pi^*(V-W+Q)$, i.e., the galois groups over K respectively of (1) $\bar\Omega(V-W+Q)$, (2) $\bar\Omega'(V-W+Q)$, (3) $\bar\Omega^*(V-W+Q)$ will be denoted respectively by (1) $\bar\pi(V-W+Q)$, (2) $\bar\pi'(V-W+Q)$, (3) $\bar\pi^*(V-W+Q)$ and they will be called respectively

(1) *the topologized fundamental group of* $V-W+Q$,

(2) *the topologized tame fundamental group of* $V-W+Q$,

(3) *the topologized reduced fundamental group of* $V-W+Q$.

Now let $W_1,\cdots,W_s$ be distinct irreducible $n-1$ dimensional sub-

varieties of V and let $Q_1, Q_2, \cdots$ be valuations of K/k of k-dimension $n-1$. Let $w_1, \cdots, w_t, q_1, q_2, \cdots$ be non-negative integers. Then we set

$$\Omega'(V-(W_1, w_1)-\cdots-(W_s, w_s)+(Q_1, q_1)+(Q_2, q_2)+\cdots)$$

$=$ the family of members L of $\Omega'(V-W_1-\cdots-W_t+Q_1+Q_2+\cdots)$ such that (1) for $j=1,2,\cdots,s$, for any local ring R^*_j in L lying above $R_j=Q(W_j, V)$ we have that $r(R^*_j : R_j)$ divides w_j; and (2) for $1,2,\cdots$, for any L-extension Q^*_j of Q_j we have that $r(Q^*_j : Q_j)$ divides q_j.

Similarly we define Ω'_g, Ω^*, Ω^*_g, $\bar{\Omega}'$, $\bar{\Omega}^*$, π', π^*, $\bar{\pi}'$ and $\bar{\pi}^*$ for $V-(W_1, w_1)-\cdots-(W_s, w_s)+(Q_1, q_1)+\cdots$. Note that in view of Lemmas 2 and 3 of Section 2, all this is justified. Also note that: (1) if Q is empty, then $\pi(V-W)=\pi(V-W+Q)$, etc.; and (2)

$$\Omega'(V-(W_1, 0)-\cdots-(W_s, 0)+(Q_1, 1)+(Q_2, 1)+\cdots)$$
$$=\Omega'(V-W_1-\cdots-W_s+Q_1+Q_2+\cdots), \text{ etc.}$$

The rest of the concepts of Section 10 of [A1] are now similarly extended.

We make here the following observation. Let K^* be a (finite) galois extension of K in Ω, let V^* be a K^*-normalization of V, let ϕ be the rational map of V^* on V and let $W^*=\phi^{-1}(W)$. Then from the results of [A1, Section 2] we deduce that if $K^* \in \Omega(V-W)$, then $\bar{\Omega}(V^*-W^*)=\bar{\Omega}(V-W)$ and if $K^* \in \Omega'(V-W)$ then $\bar{\Omega}'(V^*-W^*)=\bar{\Omega}'(V-W)$. Also [see the proof of Lemma 8 in (13) of Remark 6 of Section 8 of A2] if K_1 is a galois extension of K, K_2 is a galois extension of K_1 and K_3 is a least galois extension of K containing K_2, and if $[K_1 : K]$ and $[K_2 : K_1]$ are prime to p, then $[K_3 : K]$ is prime to p. From this it follows that if $K^* \in \Omega^*(V-W)$, then $\bar{\Omega}^*(V^*-W^*)=\bar{\Omega}^*(V-W)$. Consequently, if K^* belongs respectively to $\Omega(V-W)$, $\Omega'(V-W)$, $\Omega^*(V-W)$, then we have the following respective natural exact sequences:

$$0 \to \bar{\pi}(V^*-W^*) \to \bar{\pi}(V-W) \to G(K^*/K) \to 0,$$
$$0 \to \bar{\pi}'(V^*-W^*) \to \bar{\pi}'(V-W) \to G(K^*/K) \to 0,$$
$$0 \to \bar{\pi}^*(V^*-W^*) \to \bar{\pi}^*(V-W) \to G(K^*/K) \to 0.$$

We would like to point out that our notation $\Omega'(V-W)$, $\bar{\Omega}'(V-W)$, $\bar{\pi}(V-W)$, etc., is not meant to suggest the open variety $V-W$ as a biregular entity; see Remark 3 in Section 4 (probably, $\Omega'(V, W)$, etc., might have been a more logical notation). For $\Omega(V-W)$, $\bar{\pi}(V-W)$, etc., and for $\Omega^*(V-W)$, $\bar{\pi}^*(V-W)$, etc., there is of course no such problem because they are obviously biregular invariants.

(F). We shall use the following notations.

Z_n — a cyclic group of finite order n.

Z — an infinite cyclic group.

Z^p — the group tower of all the factor groups of Z of finite order prime to p.

$\bar{Z}^p$ — the inverse limit of Z^p.

Let K be a field extension of k and let Ω be an algebraic closure of K. Let x be an element of K. We set $\Omega(K,x)$ = the field generated over K by the n-th roots of x in Ω for all n prime to p. It is clear that if x does not have any n-th root in K for any integer $n > 1$ then $G(\Omega(K,x)/K)$ is isomorphic to $\bar{Z}^p$. Note that for any positive integer n prime to p, $\bar{Z}^p$ has a unique subgroup Y_n of index n and furthermore $\bar{Z}^p/Y_n \approx Z_n$; namely, Y_n corresponds to $G(\Omega(K,x)/K(x^{1/n}))$.

(G). Let $G, G_1, \cdots, G_s, H_1, \cdots, H_s$ be groups. The direct sum of $G_1, \cdots, G_s$ will be denoted by $G_1 \oplus \cdots \oplus G_s$. Similarly for homomorphisms f_i: $G_i \to H_i$, we shall denote by $f_1 \oplus \cdots \oplus f_s$ the homomorphism

$$G_1 \oplus \cdots \oplus G_s \to H_1 \oplus \cdots \oplus H_s$$

given by

$$(f_1 \oplus \cdots \oplus f_s)(g_1 \oplus \cdots \oplus g_s) = f_1(g_1) \oplus \cdots \oplus f_s(g_s), \quad \text{for all } g_i \text{ in } G_i.$$

For homomorphisms q_i: $G \to H_i$ we shall denote by $(q_1, \cdots, q_s)$ the homomorphism $G \to H_1 \oplus \cdots \oplus H_s$ given by

$$(q_1, \cdots, q_s)(g) = q_1(g) \oplus \cdots \oplus q_1(g), \text{ for all } g \text{ in } G.$$

The set of elements of $G \oplus G$ of the form $g \oplus g$ with g in G, forms a subgroup of $G \oplus G$ and we shall call it *the diagonal of* $G \oplus G$ and we shall denote it by $D(G \oplus G)$. Note that, if G is an additive abelian group, then $g_1 \oplus g_2 \to g_1 - g_2$ is an epimorphism of $G \oplus G$ onto G whose kernel is $D(G \oplus G)$.

2. Preliminary lemmas. For the purposes of this and succeeding papers of the series, in this section we shall prove some lemmas mainly on local theory. Modifying the proof of Lemma 1 of [A3] we get Lemma 1 below under which the quoted lemma is subsumed. Using the method of removing

tame ramification through cyclic compositums [A3, Proposition 1] we deduce Lemmas 2 and 3 below, and Proposition 7 of Section 5 of [A5] is subsumed under Lemma 3 below.

LEMMA 1. *Let (R, M) be either the quotient ring of an irreducible subvariety of a normal irreducible algebraic variety or the valuation ring of a real discrete valuation. Let K be the quotient field of R, let L^* be a finite separable extension of K and let L and K^* be two fields between K and L^* such that L^* is the K-compositum of K^* and L. Let S^* be a local ring in L^* lying above R, let $S = L \cap S^*$, and let $R^* = K^* \cap S^*$. Let D, D^*, E, E^* be the residue fields respectively of R, R^*, S, S^*, where D, D^*, E are canonically considered to be subfields of E^*. Assume that S is unramified over R. Then (I) E^* is the compositum of D^* and E; (II) S^* is unramified over R^*; and (III) $r(S^*: S) = r(R^*: R)$, $\bar{r}(S^*: S) = \bar{r}(R^*: R)$, $i(S^*: S) = i(R^*: R)$, and $g(S^*: S) \leqq g(R^*: R)$. Furthermore, (IV) if either E/D is galois (which is the case if L/K is galois) and D^* is D-isomorphic to a subfield of E or if D^*/D is galois (which is the case if K^*/K is galois) and E is D-isomorphic to a subfield of D^*, then $d(S^*: R^*) = 1$.*

Proof. Since the situation remains parallel if we pass to the completions of R, S, R^*, S^*, we may assume that these local domains are complete to begin with. Then our assumption implies that $[E: D] = [E: D]_s = [L: K]$. Since E/D is separable, there exists a in E such that $E = D(a)$. Fix an element A in S belonging to the residue class a. Let $F(X)$ be the minimal monic polynomnal of A over K. Then $F(X) \in R[X]$ and $L = K(A)$. Let $f(X)$ be the polynomial obtained by reducing the coefficients of $F(X)$ modulo the maximal ideal in R. Then $\Delta(F)$ belongs to the residue class $\Delta(f)$ and $\Delta(f) \neq 0$ since a is separable over D. Therefore $\Delta(F)$ is a unit in R and hence *$1, A, A^2, \cdots, A^{n-1}$ is a module basis of S over R, where $n = [L: K]$.* Since L^* is the K-compositum of K^* and $L, L^* = K^*(A)$. Let $G(X)$ be the minimal monic polynomial of A over K^*. Then $F(X) = G(X)H(X)$ and $G(X), H(X) \in R^*[X]$. Consequently $\Delta(G)$ divides $\Delta(F)$ in R^*. Since $\Delta(F)$ is a unit in R, it is also a unit in R^*. Therefore $\Delta(G)$ is a unit in R^*. Consequently S^* is unramified over R^*, *and $1, A, A^2, \cdots, A^{m-1}$ is a module basis of S^* over R^*, where $m = [L^*: K^*] \leqq n$.* The two italicized statements imply (I), (II) has already been proved, and (III) follows from (II) as in the proof of Lemma 1 of [A3]. Finally, (IV) follows from (I) and (II) by galois theory.

LEMMA 2. *Let K be a field, let K^* be a finite separable extension of K,*

and let $K_1, K_2, \cdots, K_s$ *be fields between* K *and* K^* *such that* K^* *is the compositum of* $K_1, \cdots, K_s$. *Let* v *be a real discrete valuation of* K, *let* v^* *be an extension of* v *to* K^* *and let* v_j *be the* K_j*-restriction of* v^*. *Let* $n^* = \bar{r}(v^*: v)$ *and* $n_j = \bar{r}(v_j: v)$. *Let* p *be the characteristic of the residue field of* v *and assume that* $i(v_1: v) = \cdots = i(v_s: v) = 1$ *and that* $n_1, \cdots, n_s$ *are prime to* p *in case* $p \neq 0$. *Then* $i(v^*: v) = 1$ *and* n^* *is the least common multiple of* $n_1, \cdots, n_s$.

Proof. Let n be the least common multiple of $n_1, \cdots, n_s$ and let x be an element of K with $v(x) = 1$. Let y be a root of the polynomial $X^n - x$ in some extension of K^*, and let $L = K(y)$, $L_j = K_j(y)$, $L^* = K^*(y)$. Let w^* be an extension of v^* to L^* and let w_j and w be the restrictions of w^* to L_j and L respectively. By Proposition 1 of [A3], $w_1, \cdots, w_s$ are unramified over w and hence by repeated applications of part (III) of Lemma 1 we conclude that w^* is unramified over w. Now $w(y) = (1/n)w(x)$ and $n = [L: K]$ and hence we must have $r(w: v) = \bar{r}(w: v) = n$. Therefore $i(w^*: v) = 1$ and $\bar{r}(w^*: v) = n$. Hence $i(v^*: v) = 1$ and $\bar{r}(v^*: v)$ divides n. Also, it is clear that each n_j divides n^* and hence n divides $\bar{r}(v^*: v)$. Therefore $n^* = n$.

LEMMA 3. *Let* K *be a field, let* K_1 *be a finite separable extension of* K *and let* K^* *be a least galois extension of* K *containing* K_1. *Let* v *be a real discrete valuation of* K, *let* $v_1, \cdots, v_s$ *be the distinct extensions of* v *to* K_1 *and let* v^* *be any extension of* v *to* K^*. *Let* $n_j = \bar{r}(v_j: v)$ *and* $n^* = \bar{r}(v^*: v)$. *Let* p *be the characteristic of the residue field of* v *and assume that* $i(v_1: v) = \cdots = i(v_s: v) = 1$ *and that* $n_1, \cdots, n_s$ $n_1, \cdots, n_s$ *are prime to* p *in case* $p \neq 0$. *Then* $i(v^*: v) = 1$ *and* n^* *is the least common multiple* $n_1, \cdots, n_s$.

Proof. Let $1, t_2, \cdots, t_h$ be the K-automorphism of K^* and let $K_j = t_j(K_1)$. Let w_j be the K_j-restriction of v^* to K_j. Then $i(w_1: v) = \cdots = i(w_h: v) = 1$ and the set of integers $\bar{r}(w_1: v), \cdots, \bar{r}(w_h: v)$ coincides with the set $n_1, \cdots, n_s$. Thus we are reduced to Lemma 2.

LEMMA 4. *Let* y *be a transcendental over* k *and let* $k(y)((x))$ *be the formal power series field over* $k(y)$ *in an indeterminate* x. *Let* $f(Z)$ *be an irreducible monic polynomial in* $k[[x]][Z]$ *of degree* n *which is prime to* p *in case* $p \neq 0$. *Let* z *be a root of* $f(Z)$ *in some extension of* $k(y)((x))$. *Let* v *and* w *be the valuations of* $k((x))$ *and* $k(y)((x))$ *with the valuation rings* $k[[x]]$ *and* $k(y)[[x]]$ *respectively. Let* v^* *and* w^* *be the (unique) extensions of* v *and* w *to* $k((x))(z)$ *and* $k(y)((x))(z)$ *respectively. Then*

$k((x))(z)/k((x))$ and $k(y)((x))(z)/k(y)((x))$ are cyclic extensions of degrees n; and $\bar{r}(v^*: v) = \bar{r}(w^*: w) = n$ and

$$i(v^*: v) = i(w^*: w) = 1 = g(v^*: v) = g(w^*: w).$$

Proof. By [A5, Lemma A5 of Section 8], $K((x))(z)$ contains an n-th root $x^{1/n}$ of x, and $k((x))(z) = k((x))(x^{1/n})$. Hence $k(y)((x))(z) = k(y)((x))(x^{1/n})$. Now everything follows in view of the equations: $v(x) = w(x) = 1$.

LEMMA 5. *Let K be an n-dimensional algebraic function field over k, let V be a normal projective model of K/k, let V^* be a normailzation of V in a finite separable extension K^* of K, let ϕ be the rational map of V^* onto V, let P^* be a point of V^* and let $P = \phi(P^*)$. Assume that P is a simple point of V, and that $\Delta(K^*/V)$ is contained in a pure $n-1$ dimensional subvariety W of V having a strong normal crossing at P such that the irreducible components $W_1, \cdots, W_t$ of W passing through P are all tamely ramified in K^*. Then* (I) *P is tamely ramified in K^*; and* (II) *W_j does not split locally at P^*, i. e., only one irreducible component W^*_j of $\phi^{-1}(W_j)$ passes through P^*.*

Now assume that $W_2, \cdots, W_t$ are unramified in K^*. Let v be a real discrete valuation of K/k of k-dimension $(n-1)$, having center P on V, let v^* be a K^*-extension of v having center P^* on V^*. Let $m = r(W^*_1: W) = \bar{r}(W^*_1: W)$, let $q = v(M(P, W_1, V))$, and let h be the greatest common divisor of m and q. Then (III) *$V^*, W^*_1, \cdots, W^*_t$ all have a simple point at P^*, and $r(P^*: P) = r(W^*_1: W_1)$;* (IV) *$r(v^*: v) = \bar{r}(v^*: v) = m/h$.* (V) *In particular, if v is the P-adic divisor of V, then v^* is ramified over v if and only if W^*_1 is ramified over W_1.* (VI) *Now assume $n = 2$ and let $\bar{W}$ be an irreducible curve on V other than W^*_1 such that $\bar{W}$ has a simple point at P and W^*_1 and W have an m-fold contact at P; then $\phi^{-1}(\bar{W})$ has exactly m irreducible components $\bar{W}^*_1, \cdots, \bar{W}^*_m$ passing through P^*; furthermore $W^*_1, \bar{W}^*_1, \cdots, \bar{W}^*_m$ all have a simple point at P^* and they are pairwise non-tangential at P^*; consequently $\phi^{-1}(W_1 \cup \bar{W})$ has an ordinary $(m+1)$-fold point at P^*.*

Proof. Let m_j be the least common multiple of the ramification indices over W_j of the various irreducible components of $\phi^{-1}(W_j)$. Let (R, M) be the quotient ring of P on V and let (R^*, M^*) be the quotient ring of P^* on V^*. Fix a basis $x_1, \cdots, x_n$ of M such that x_jR is the ideal of W_j at P on V for $j = 1, \cdots, t$. Let $m_j = 1$ for $j = t+1, t+2, \cdots, n$. Let y_j be an m_j-th root of x_j in some extension of K^*. Let $L^* = K^*(y_1, \cdots, y_n)$ and $L = K(y_1, \cdots, y_n)$. Let (S^*, N^*) be a local ring in L^* lying above R^*

and let (S,N) be the local ring in L lying below (S^*,N^*). Let $(\bar{R},\bar{M})$, $(\bar{R}^*,\bar{M}^*)$, $(\bar{S},\bar{N})$, $(\bar{S}^*,\bar{N}^*)$ be the completions of R, R^*, S, S^* respectively and let E, E^*, F, F^* be the quotient fields of $\bar{R}$, $\bar{R}^*$, $\bar{S}$, $\bar{S}^*$ respectively. Then $y_1,\cdots,y_n$ is a basis of $\bar{N}$ and hence $\bar{S}$ and S are regular and $y_1,\cdots,y_n$ is a basis of N. By Proposition 1 of [A3], no one-dimensional prime ideal in S is ramified in L^* and hence by [Z], S^* is unramified over S. Consequently, $y_1,\cdots,y_n$ is a basis of N^* and we have

$$k((x_1,\cdots,x_n))=E\subset E^*\subset F^*=k((y_1,\cdots,y_n)).$$

Therefore E^*/E is galois and its degree is prime to p in case $p\neq 0$. Since P^* was an arbitrary point of $\phi^{-1}(P)$, this proves (I). Since $y_j^{m_j}=x_j$ and y_j is part of a minimal basis of N^*, x_jS^* is primary. Therefore the valuation of K/k with valuation ring $Q(W_j,V)$ has a unique extension to L^* having center N^* in S^* and consequently the said valuation *a fortiori* has a unique extension to K^* having center M^* in R^*, i.e., $\phi^{-1}(W_j)$ has only one component W^*_j passing through P^*. This proves (II).

Now assume that $W_2,\cdots,W_t$ are unramified in K^*. Then $m_2=m_3=\cdots=m_n=1$. Therefore

$$k((x_1,\cdots,x_n))=E\subset E^*\subset F^*=k((y_1,x_2,\cdots,x_n)).$$

Consequently $E^*=k((x_1^{1/u},x_2,\cdots,x_n))$, where u is an integer which is prime to p in case $p\neq 0$. From this it follows that $V^*,W^*_2,\cdots,W^*_n$ have a simple point at P^*. Since P^* is a simple point of V^*, R^* is a unique factorization domain, $M(P^*,W^*_1,V^*)$ is a principal ideal in R^* and $x_1=z^md$, where z is a generator of $M(P^*,W^*_1,V^*)$, a is a positive integer and d is a unit in R^*. By a well-known property of quotient rings on a normal algebraic variety, the completion of $R^*/(zR^*)$ has no nonzero nilpotents and hence we must have $u=m$ and $z=x_1^{1/u}e$, where e is a unit in $\bar{R}^*$. Therefore W^*_1 also has a simple point at P^*. This proves (III). Let $\bar{v}^*$ be an extension of v^* to E^* having center $\bar{M}^*$ in $\bar{R}^*$ [A6, Lemma 13 of Section 7] and let $\bar{v}$ be the E-restriction of $\bar{v}^*$. Then $\bar{v}$ is an extension of v to E and has center $\bar{M}$ in $\bar{R}$. Now $\bar{v}^*$ and $\bar{v}$ are again real discrete [A4, Theorem 1] and have the same value groups and residue fields as v^* and v respectively [A6, Lemma 12 of Section 7]. Hence in the proof of (IV) we may replace v and v^* by $\bar{v}$ and $\bar{v}^*$ respectively. Now $\bar{v}(x_1)=v(x_1)=q$ and $E^*=E(x_1^{1/m})$ and hence (IV) follows by Lemma 3 of [A3]. In case v is the P-adic divisor of V, we have $q=1$ and this gives (V).

Finally, assume that the conditions of (VI) are satisfied. Let $x=x_1$. We can choose w in M such that (x,w) is a basis of M and wR is nontangential

at P to W_1 as well as to $\bar{W}$. By [A2, Section 7], there exists elements $a_0, a_1, \cdots$ in k with $a_0 \neq 0$ such that $w^* = x - (a_0 w^m + a_1{}^{m+1} + \cdots)$ is a generator of $M(P, \bar{W}, V)$. As in the above proof, (y, w) is a minimal basis of $\bar{M}^*$, where $y^m = x$. Now the $\bar{R}^*$-leading form of w^* is $y^m - a_0 w^m$. Since m is prime to p in case $p \neq 0$, this leading form factors into m pairwise co-prime linear factors each of which is prime to y. This proves (VI).

COROLLARY. *Let V be a normal projective irreducible algebraic variety over k, let W be an irreducible subvariety of V of codimension one, and let K be a finite algebraic extension of $k(V)$. If there exists a point of P of W which is tamely ramified in K, then all points of W lying outside a proper subvariety of W are tamely ramified in K.*

Our assumption implies that [A1, Section 2] $Q(W, V)$ is tamely ramified in K. Choose a subvariety W^* of V of pure codimension one such that $\Delta(K/V) \subset W^*$ and $W \subset W^*$. Then by Lemma 5 every point W which is a simple point of W as well as a simple point of V is tamely ramified in K.

LEMMA 6. *Let K be an n-dimensional algebraic function field over k, let K^* be a finite separable extension of K, let V and V' be two normal projective models of K/k and let f be the birational map of V onto V'. Assume that* (1) *any point of V at which f is not regular is a simple point of V;* (2) *$f^{-1}(\Delta(K^*/V')) \subset W$, where W is a pure $n-1$ dimensional subvariety of V having a strong normal crossing at every point of V which is not regular for f; and* (3) *K^*/V' is tamely ramified. Then K^*/V is tamely ramified.*

Proof. Without loss of generality, we may assume that K^*/K is galois. Let P be a point of V. If f is regular at P, then in view of [A6, Lemma 6 of Section 2], (3) implies that P is tamely ramified in K^*. Now assume that f is not regular at P. If P is unramified in K^*, then there is nothing to prove, so assume that P is ramified in K^*. Then [Z] every irreducible component A of $\Delta(K^*/V)$ passing through P is $(n-1)$-dimensional; let w be the real discrete valuation of K/k having center A on V and let A' be the center of w on V'; then by [A6, Lemma 6 of Section 2] w is tamely ramified in K^* and $A' \subset \Delta(K^*/V')$ and hence A is a component of W. Consequently, in view of (2), it follows that $\Delta(K^*/V) \subset B$, where B is a pure $(n-1)$-dimensional subvariety of V such that B has a strong normal crossing at P and such that every irreducible component of B passing through P is tamely ramified in K^*. Therefore by Part I of Lemma 5, P is tamely ramified in K^*.

Remark 1. Let K be an n-dimensional algebraic function field over k, let V and V' be two normal models of K/k, and let f be the map of V onto V',

and let K^* be a finite separable extension of K. One might raise the following questions concerning birational invariance of tame ramification. (1) (*Birational invariance*): Is it true that if K^*/V' is tamely ramified, then K^*/V is necessarily tamely ramified? (2) (*Birational nonsingular invariance*): Question (1) under the assumption that V and V' are nonsingular. Note that there exist proper subvarieties W and W' of V and V' respectively such that f is biregular between $V-W$ and $V-W'$, $V-W$ and $V-W'$ are nonsingular and each point of them is unramified in K^*. An affirmative answer to (1) would have had the above lemma as well as [A2, Lemma 6 of Section 2] as immediate corollaries. However, as will be seen in Remark 3 of Section 4, in general (2) and hence (1) have a negative answer.

LEMMA 7. *Let K be a field, let K^* be a galois extension (not necessarily finite) of K, let K_1 and K_2 be fields between K and K^* such that K_1/K and K_2/K are galois (not necessarily finite) and K^* is the compositum of K_1 and K_2. Let $K'=K_1\cap K_2$ and assume that $G(K'/K)$ is abelian. Then for $j=1,2$, $G(K_j/K)$ contains a normal subgroup G_j such that $G(K_j/K)/G_j \approx G(K'/K)$ and there exists an exact sequence*

$$0\to G(K^*/K)\xrightarrow{f} G(K_1/K)\oplus G(K_2/K)\to G(K'/K)\to 0,$$

such that $G_1\oplus G_2\subset(\operatorname{Im} f)$.

Proof. The first statment follows by taking $G_j=G(K_j/K')$. Now consider the natural exact sequences

$$(1)\qquad 0\to G(K^*/K_j)\to G(K^*/K)\xrightarrow{f_j} G(K_j/K)\to 0,\qquad j=1,2,$$

where $G(K^*/K_j)$ is considered to be a subgroup of $G(K^*/K)$. Now $G(K^*/K_1)G(K^*/K_2)=G(K^*/K')$ and hence

$$(2)\qquad (f_1,f_2)G(K^*/K')=G_1\oplus G_2\subset(f_1,f_2)G(K^*/K).$$

From (1) and (2) we get the exact sequence

$$0\to(f_1,f_2)G(K^*/K')\to G(K_1/K)\oplus G(K_2/K)\xrightarrow{g} G(K'/K)\oplus G(K'/K)\to 0,$$

where g is the natural epimorphism and we have that

$$(g(f_1,f_2))G(K^*/K)=D(G(K'/K)\oplus G(K'/K)).$$

Since $G(K'/K)$ is abelian, we have an exact sequence

$$0\to D(G(K'/K)\oplus G(K'/K))\to G(K'/K)\oplus G(K'/K)\xrightarrow{h} G(K'/K)\to 0.$$

Thus we get the exact sequence

$$0 \to G(K^*/K) \xrightarrow{(f_1, f_2)} G(K_1/K) \oplus G(K_2/K) \xrightarrow{hg} G(K'/K) \to 0.$$

LEMMA 8. *Let K be a field; let Ω be an algebraic closure of K; let $A_1, \cdots, A_t$ $(t > 1)$ be abelian galois extensions (not necessarily finite) of K in Ω; let B_j be the compositum of $A_1, \cdots, A_j$ in Ω. Then there exists an exact sequence*

$$0 \to G(B_t/K) \to G(A_1/K) \oplus \cdots \oplus G(A_t/K) \to F \to 0$$

where F is an abelian group which contains a chain of subgroups $F = F_1 \supset F_2 \supset \cdots \supset F_t = 1$ such that $F_{j-1}/F_j \approx G((B_{j-1} \cap A_j)/K)$ for $j = 2, \cdots, t$.

Proof. Taking A_1 for K_1 and A_2 for K_2 in Lemma 7 we get the case of $t = 2$. Since all the groups are abelian, the case of $t > 2$ follows from the case of $t - 1$ by taking B_{t-1} for K_1 and A_t for K_2 in Lemma 7. Thus our assertion follows by induction on t.

3. A variety minus two sets of irreducible subvarieties of codimension one.

Let V be a nonsingular projective irreducible algebraic variety over k; let $W_1, \cdots, W_s$, $H_1, \cdots, H_t$ be distinct irreducible subvarieties of V of codimension one; let $W = W_1 \cup \cdots \cup W_s$ and $H = H_1 \cup \cdots \cup H_t$; let $K = k(V)$; and let Ω be an algebraic closure of K. As in Proposition 2 of [A3], using the technique of removing tame ramification through cyclic compositums and modifying the proof of that proposition, we now prove the following results.

PROPOSITION 1. *Assume that $Q(H_1, V)$ does not split*[2] *in any member of $\Omega'_g(V - W - H_1)$ and that $\alpha H_1 \equiv 0 \pmod{W_1, \cdots, W_s}$ for some nonzero integer α. Choose the smallest positive value of α and let a be the part of α prime to p. Fix x in K such that $(x) - H_1$ equals a sum of integral multiples of $W_1, \cdots, W_s$. Then* (I) *x does not have any m-th root in K for any $m > 1$;* (II) *$\bar{\Omega}'(V - W - H_1)$ is the compositum of $\bar{\Omega}'(V - W)$ and $\Omega(K, x)$;*

and (III) $$\Omega'(V - W) \cap \Omega(K, x) = K(x^{1/a}).$$

Proof. (I) follows from the minimality of α. Also, it is obvious that the compositum of $\bar{\Omega}'(V - W)$ and $\Omega(K, x)$ is contained in $\bar{\Omega}'(V - W - H_1)$. Now we shall show that any member K^* of $\Omega'(V - W - H_1)$ is contained in

[2] I. e., in any member of $\Omega'_g(V - W - H_1)$ there is only one local ring lying above $Q(H_1, V)$.

the said compositum. Without loss of generality we may assume that K^*/K is galois. Let $R = Q(H_1, V)$, let R^* be the unique local ring in K^* lying above R, and let $m = r(R^* : R)$. Let y be an (ma)-th root of x in Ω. Let $L = K(y)$ and $L^* = K^*(y)$. Then L^* belongs to $\Omega'_g(V - W - H_1)$. Let S and S^* be the unique local rings lying above R respectively in L and L^*. By [A3, Lemma 3], $r(S: R) = m$. Hence by Lemma 2, $r(S^*: R) = m$. Let K' be the inertia field of S^* over R and let L' be the compositum of K' and L. Let R' and S' be the unique local rings lying above R respectively in K' and L'. Then K' belong to $\Omega'_g(V - W)$. Also

$$[L' : K'] \leqq [L^* : K'] = r(S^* : R) = m; \text{ and}$$
$$[L' : K'] \geqq r(S' : R') = r(S' : R) \geqq r(S : R) = m.$$

Therefore $L^* = L'$. This proves (II). Now let b be any positive integer prime to p. Then [A3, Proposition 1], R is unramified in $K(x^{1/b})$ if and only if b divides a. This proves (III).

THEOREM 1. *Assume that*: (I) *the subvarieties H_j can be labelled so that for $j = 1, \cdots, t$; $Q(H_j, V)$ does not split in any member of $\Omega'_g(V - W - H_1 - \cdots - H_j)$. Also assume that*: (II) *for each j, some nonzero integral multiple of H_j is linearly equivalent to zero* mod $W_1, \cdots, W_s$. *Let α_j be the smallest positive integer such that $\alpha_j H_j \equiv 0$ (mod $W_1, \cdots, W_s$). Let a_j be the part of α_j prime to p. Then there exists an exact sequence*

$$0 \to \bar{\pi}'(V - W - H) \to \bar{\pi}'(V - W) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \to G \to 0$$

where $\bar{Z}^p$ occurs t times, and G is an abelian group containing a sequence of subgroups $G = G_1 \supset G_2 \supset \cdots \supset G_t$ such that G_t is isomorphic to a subgroup of $Z_{a_1} \oplus \cdots Z_{a_t}$; and G_{j-1}/G_j is isomorphic to a subgroup of Z_{a_j} for $j = 2, \cdots, t$. If $t = 1$ then the above exast sequence can be chosen so that $G \approx Z_{a_1}$.

Proof. Fix x_j in K such that $(x_j) = \alpha_j H_j +$ integral multiples of $W_1, \cdots, W_s$. Fix $y_j \in \Omega$ such that $y_j^{\rho_j} = x_j$. Let $K'(y_1, \cdots, y_t)$. Let $A_j = \Omega(K, x_j)$, and $B_j =$ (compositum of $A_1, \cdots, A_j$). Let $C = \bar{\Omega}'(V - W)$, and $D = \bar{\Omega}'(V - W - H)$. Then $\bar{\pi}'(V - W) = G(C/K)$, and $\bar{\pi}'(V - W - H) = G(D/K)$. By Proposition 1, $D =$ compositum of C and B_t, and $K(y_1, \cdots, y_t) \subset C \cap B_t$. Suppose if possible that $K' \neq C \cap B_t$. Then there exists a cyclic extension L of K' of prime degree $q \neq p$ such that $L \subset C \cap B_t$. We can find a primitive element z of L/K' such that $z^q = y_1^{b_1} \cdots y_t^{b_t}$. Then q does not divide b_j for some j and hence H_j is ramified in L. This contradicts the fact that $L \subset C$. Therefore $C \cap B_t = K'$. The minimality of α_j

implies that K contains no m-th root of x_j for any integer $m > 1$. Hence $G(K(y_j)/K) \approx Z_{a_j}$, and $G(A_j/K) \approx \bar{Z}^p$ for $j = 1, \cdots, t$.

Taking C for K_1 and B_t for K_2 in Lemma 7 we get the following: $G(C/K)$ and $G(B_t/K)$ contain normal subgroups X and Y respectively, such that $G(C/K)/X$ is abelian, and there exists an exact sequence

$$0 \to G(D/K) \xrightarrow{u} G(C/K) \oplus G(B_t/K) \to G(K'/K) \to 0$$

such that $X \oplus Y \subset \operatorname{Im} u$.

By [A3, Proposition 1], $B_{j-1} \cap A_j \subset K(y_j)$ and hence $G((B_{j-1} \cap A_j)/K)$ is isomorphic to a subgroup of Z_{a_j}. Hence by Lemma 8 we get an exact sequence

$$0 \to G(B_t/K) \xrightarrow{v} \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \to F \to 0$$

where $\bar{Z}^p$ occurs t times and F is an abelian group containing a chain of subgroups $F = F_1 \supset F_2 \supset \cdots \supset F_t = 1$ such that F_{j-1}/F_j is isomorphic to a subgroup of Z_{a_j} for $j = 2, \cdots, t$.

Let i: $G(C/K) \to G(C/K)$ be the identity map. Let $w = (i \oplus v)u$. Now X is a normal subgroup of $G(C/K)$, $G(C/K)/X$ is abelian, and $\bar{Z}^p \oplus \cdots \oplus \bar{Z}^p$ is abelian. Therefore $X \oplus vY$ is a normal subgroup of $G(C/K) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p$ and the corresponding factor group is abelian. Since $X \oplus Y \subset \operatorname{Im} u$, we get that $X \oplus vY \subset \operatorname{Im} w$. Therefore, $\operatorname{Im} w$ is a normal subgroup of $G(C/K) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p$ and letting

$$G = [G(C/K) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p]/\operatorname{Im} w$$

we get that G is abelian. Let f: $G(C/K) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \to G$ be the natural epimorphism. Now it follows that

$$0 \to G(D/K) \xrightarrow{w} G(C/K) \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \xrightarrow{f} G \to 0$$

is exact, G is an abelian group containing a sequence of subgroups $G = G_1 \supset G_2 \supset \cdots \supset G_t$ such that $G_t \approx G(K'/K)$, and $G_{j-1}/G_j \approx F_{j-1}/F_j$ for $j = 2, \cdots, t$.

Taking $K(y_j)$ for A_j in Lemma 8, we deduce that $G(K'/K)$ is isomorphic to a subgroup of $Z_{a_1} \oplus \cdots \oplus Z_{a_t}$.

Finally, in case of $t = 1$ it is obvious that $G(B_t/K) \approx \bar{Z}^p$ and $G(K'/K) \approx Z_{a_1}$.

COROLLARY 1. *Theorem* 1 *holds if condition* (I) *is replaced by the condition* (IA) : *$W \cup H$ has a strong normal crossing at every point of H and* $\dim |H_j| > 1$ *for* $j = 1, \cdots, t$.

Follows from Theorem 1, in view of [A1, proof of Proposition 6 of Section 11].[3]

COROLLARY 2. *Theorem* 1 *holds if condition* (I) *is replaced by condition* (IB) : $n=2$ *and for some labelling of the curves* H_j *we have*

$$\dim |H_j| > 1 + \nu(H_j, W \cup H_1 \cup \cdots \cup H_j) \quad for \ j=1,\cdots,t.$$

Proof. Follows from from Theorem 1, in view of [A2, Proposition 14 of Section 9].

COROLLARY 3. *Assume that the following conditions hold:* (I) W *has a strong normal crossing at every point of* V *and* $\dim |W_j| > 1$ *for* $j=1,\cdots,s$; (II) *some nonzero integral multiple of* W_j *is linearly equivalent to some integral multiple of* W_k *whenever* $j \neq k$; (III) V *is simply connected. Let* α_j *be the smallest positive integer such that* $\alpha_j W_j \equiv 0 \pmod{W_1}$, $(j>1)$. *Let* a_j *be the part of* α_j *prime to* p, *and let* $a=\delta(W_1, V)$. *Then there exists an exact sequence*

$$0 \to \bar{\pi}'(V-W) \to Z_a \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p \to G \to 0,$$

where $\bar{Z}^p$ *occurs* $s-1$ *times and* G *is an abelian group containing a sequence of subgroups* $G=G_2 \supset G_3 \supset \cdots \supset G_s$ *such that* G_s *is isomorphic to a subgroup of* $Z_{a_2} \oplus \cdots \oplus Z_{a_s}$, *and* G_{j-1}/G_j *is isomorphic to a subgroup of* Z_{a_j} *for* $j=3,\cdots,s$. *Consequently* $\bar{\pi}'(V-W)$ *is abelian and* $\bar{\pi}'(V-W) = \bar{\pi}^*(V-W)$. *Furthermore, if the subvarieties* W_j *can be labelled so that* $W_j \equiv 0 \pmod{W_1}$ *for* $j=2,\cdots,s$, then

$$\bar{\pi}'(V-W) \approx Z_a \oplus \bar{Z}^p \oplus \cdots \oplus \bar{Z}^p,$$

where $\bar{Z}^p$ *occurs* $s-1$ *times and again* $a=\delta(W_1, V)$.

Follows from Corollary 1 and [A1, Theorem 4 of Section 14][3].

COROLLARY 4. *Corollary* 3 *holds if the assumption* (I) *is replaced by assumption* (IA) : $n=2$ *and for some labelling of the curves* W_j *we have*

$$\dim |W_j| > 1 + \nu(W_j, W_1 \cup \cdots \cup W_j; V) \quad for \ j=1,\cdots,t;$$

and provided the labellings used in Corollary 3 *refer to this labelling.*

Follows from Corollary 2 and [A2, Theorem 4 of Section 10].

Remark 2. Theorem 1 of [A3] is now subsumed under Corollaries 3 and 4 above. Also Proposition 2 of [A3] now follows from Proposition 1 above.

4. Projective plane minus lines. Let P^2 be a projective plane over k, let $L_1,\cdots,L$ $(s>0)$ be s distinct lines in P^2 passing thropgh a common

[3] See the correction to [A1] given in [A2, Remark 6 of Section 8].

point P and let P^1 be a line in P^2 not passing through P. Then $P_1 = L_1 \cap P^1$, $\cdots, P_s = L_s \cap P^1$ are distinct points on P^1. In a natural way, we may identify $k(P^1)$ with a subfield of $k(P^2)$. For instance: Choose projective coordinates X, Y, Z in P^2 such that $X = Z = 0$ is P, $Y = Z = 0$ is P_1, and $X = Y = 0$ is on P^1. Let $x = X/Z$ and $y = Y/Z$. Then (x, y) are affine coordinates in P^2, L_1 is the line at infinity, P^1 is the x-axis, and $L_2, \cdots, L_s$ are finite lines parallel to the y-axis. Now we can take $k(P^2) = k(x, y)$ and $k(P^1) = k(x)$.

Let Ω be an algebraic closure of $k(P^2)$ and consider extensions of $k(P^2)$ and $k(P^1)$ only in Ω.[4]

PROPOSITION 2. *Let L be a member of $\Omega'(P^1 - P_1 - \cdots - P_s)$ and let $K = L(y) =$ compositum of L and $k(P^2)$ in Ω. Then $\Delta(K/P^2) \subset L_1 \cup \cdots \cup L_s$. Let V be a K-normalization of P^2 and let ϕ be the rational map of V onto P^2. Then $W = \phi^{-1}(P^1)$ is a nonsingular irreducible curve. Let ψ be the restriction of ϕ to W. Then W is a covering of P^1 with the covering map ψ and we can consider L to be $k(W)$. $[L : k(P^1)] = [K : k(P^2)]$ and $L/k(P^1)$ is galois if and only if $K/k(P^2)$ is galois and in that case $G(L/k(P^1))$ and $G(K/k(P^2))$ are naturally isomorphic. Let $P_{j1}, \cdots, P_{jq_j}$ be the distinct point in $\psi^{-1}(P_j)$. Through each P_{jt} passes only one irreducible component L_{jt} of $\phi^{-1}(L_j)$ and every irreducible component of $\phi^{-1}(L_j)$ passes through one of the points P_{jt}. Thus $L_{j1}, \cdots, L_{jq_j}$ are exactly all the distinct irreducible components of $\phi^{-1}(L_j)$. Let $R_j = Q(L_j, P^2)$, $R_{jt} = Q(L_{jt}, V)$, $S_j = Q(P_j, P^1)$, $S_{jt} = Q(P_{jt}, W)$. Let E_j, E_{jt}, F_j, F_{jt} be the quotient fields of the completions respectively of R_j, R_{jt}, S_j, S_{jt}. Then for all values of j and t we have*

(i)
$$\begin{aligned} d(R_{jt} : R_j) &= r(R_{jt} : R_j) = \bar{r}(R_{jt} : R_j) \\ &= d(S_{jt} : S_j) = r(S_{jt} : S_j) = \bar{r}(S_{jt} : S_j)\,; \end{aligned}$$

(ii)
$$\begin{aligned} g(R_{jt} : R_j) &= i(R_{jt} : R_j) \\ &= g(S_{jt} : S_j) = i(S_{jt} : S_j) = 1\,; \text{ and} \end{aligned}$$

(iii) *E_{jt}/E_j and F_{jt}/F_j are galois extensions having a cyclic group of order $r(S_{jt} : S_j)$ for galois group.*

Now let P^*_j be any point of L_j different from P. Then L_{jt} contains exactly one point P^*_{jt} of $\phi^{-1}(P^*_j)$ and each point of $\phi^{-1}(P^*_j)$ lies on only one L_{jt}. Furthermore, (i), (ii), (iii) *hold with P_{jt} replaced by P^*_{jt} and P_j replaced by P^*_j.*

---

[4] Let S be an algebraic closure of $k(x)$ in Ω. For convenience we shall denote $S'(P^1 - P_1 - \cdots - P_s)$, $\bar{S}'(P^1 - P_1 - \cdots - P_s)$, $\cdots$, etc., respectively by

$$\Omega'(P^1 - P_1 - \cdots - P_s),\ \bar{\Omega}'(P^1 - P_1, \cdots, P_s), \cdots, \text{ etc.}$$

Every point of P^2 other than P is tamely ramified in $k(P^2)$ and V is nonsingular outside $\phi^{-1}(P)$. The P-adic divisor v of P^2 has a unique extension w to K; $g(w:v)=[k(V):k(P^2)]$, and hence v is unramified in $k(P^2)$ and $\phi^{-1}(P)$ consists of a single point. Furthermore, K/P^2 is tamely ramified if and only if $L\in\Omega^*(P^1-P_1-\cdots-P_s)$.

Let V^* be an immediate quadratic transform of P^2 with center at P. Then K/V^* is tamely ramified.

Proof.[5] First note that $k(x)$ and $k(y)$ are linearly disjoint over k and hence $[K:k(x,y)]=[L:k(x)]$ and $K/k(x,y)$ is galois if and only if $L/k(x)$ is galois and then the two galois groups are isomorphic.

$k(x)$ is a subfield of $Q(P^1,P^2)$ and under the natural epimorphism $Q(P^1,P^2)\to Q(P^1,P^2)/M(P^1,P^2)$, $k(x)$ is mapped isomorphically onto $Q(P^1,P^2)/M(P^1,P^2)$; hence after an identification, we have a natural epimorphism $\alpha\colon Q(P^1,P^2)\to k(x)$. Let W be an irreducible component of $\phi^{-1}(P^1)$. Since L is algebraic over $k(x)$, we must have $L\subset Q(W,V)$. Hence

$$[L:k(x)]=[k(V):k(P^2)]\geqq g(Q(W,V):Q(P^1,P^2))\geqq[L:k(x))].$$

Therefore $\phi^{-1}(P^1)$ is irreducible, i.e., $W=\phi^{-1}(P^1)$, and P^1 is unramified in $k(V)$; and under the natural epimorphism $Q(W,V)\to Q(W,V)/M(W,V)$, L is mapped isomorphically onto $Q(W.V)/M(W,V)$. Hence after an identification we have a natural epimorphism $Q(W,V)\to L$. It is clear that this epimorphism is an extension of α and we shall continue to denote it also by α. It is via α that we may regard $k(x)$ and L, respectively, to be $k(P^1)$ and $k(W)$. Let us note that for any point Q of W we now have:

$$\alpha(Q(Q,V))=Q(Q,W)\ \text{ and }\ Q(M(Q,V))=M(Q,W).$$

Let A be the integral closure of $k[x,y]$ in K, and let B be the integral closure of $k[x]$ in L. Then $B\subset A$ and hence via α, we conclude that every point of W lying above a point of P^1 at finite distance is a simple point of W. Also, replacing x by $1/x$ and y by y/x we can similarly conclude that every point of W lying above the point at infinity on P^1 is a simple point of W. Thus W is nonsingular and hence the restriction ψ of ϕ to W is a covering map of W onto P^1.

For $j=2,\cdots,s$, let a_j be the element of k such that the point P_j of P^2 has affine coordinates $(x=a_j,\ y=0)$, i.e., P_j as a point of P^1 has affine coordinate $(x=a_j)$. Let I and J be the ideals respectively in $k[x,y]$ and $k[x]$ generated by $(x-a_2),\cdots,(x-a_s)$. Then $J^m\subset\Delta(B/k[x])$ for some posi-

[5] We shall use the results of [A1, Section 2] and [A6, Section 2] without explicit references.

tive integer m.[6] It is clear that if $w_1, \cdots, w_n$ is a $k(x)$-basis of L belonging to B, then it is also a $k(x,y)$-basis of K belonging to A and the $L/k(x)$-discriminant of $w_1, \cdots, w_n$ coincides with its $K/k(x,y)$-discriminant. Therefore $\Delta(B/k[x]) \subset \Delta(A/k[x,y])$ and hence $I^m \subset \Delta(A/k[x,y])$. Consequently, the finite part of $\Delta(K/P^2)$ is contained in $L_2 \cup \cdots \cup L_s$ and hence $\Delta(K/P^2) \subset L_1 \cup \cdots \cup L_s$.

Let $x_j = x - a_j$. Then $F_j = k((x_j))$ and $E_j = k(x)((x_j))$. Let $b_{j1}, \cdots, b_{jq_j}$ be distinct elements of k. By the theorem of independence of valuations, there exists a primitive element z_j of $L/k(x)$ belonging to B such that

$$z_j \equiv b_{jt} (\operatorname{mod} M(P_{jt}, W)) \text{ for } t = 1, \cdots, q_j.$$

Via α, we deduce that

$$z_j \equiv b_{jt} (\operatorname{mod} M(P_{jt}, V)) \text{ for } t = 1, \cdots, q_j.$$

Let $f_j(Z)$ be the minimal monic polynomial of z_j over $k(x)$. Then

$$f_j(Z) = f_{j1}(Z) \cdots f_{jq_j}(Z), \tag{1}$$

where $f_{j1}, \cdots, f_{jq_j}$ are monic pairwise coprime irreducible polynomials in $k[[x_j]][Z]$ such that

$$f_{jt}(Z) \equiv (Z - b_{jt})^{d(S_{jt}:S_j)} (\operatorname{mod} x_j) \text{ for } t = 1, \cdots, q_j. \tag{2}$$

In view of Lemma 4, we conclude that L_j is tamely ramified in K. Hence by Lemma 5, for $t = 1, \cdots, q_j$, exactly one irreducible component L_{jt} of $\phi^{-1}(L_j)$ passes through P_{jt}. Also from (2) we deduce that $z_j \equiv b_{jh} (\operatorname{mod} M(L_{j1}, V))$ for some h. Since $A \cap M(P_{j1}, V) \supset A \cap M(L_{j1}, V)$, we have

$$z_j \equiv b_{jh} (\operatorname{mod} M(P_{j1}, V)).$$

Therefore we must have $h = 1$. Similarly, $z_j \equiv b_{jt} (\operatorname{mod} M(L_{jt}, V))$ for $t = 1, \cdots, q_j$. From this it follows that $L_{j1}, \cdots, L_{jq_j}$ are all distinct and hence they are exactly the irreducible components of $\phi^{-1}(L_j)$. Replacing x_j by $1/x$ and y by y/x, we can deal with the case of P_1 in a similar manner. In view of these considerations, (i), (ii), (iii) now follow from Lemma 4. Also in view of Lemma 5, it follows that (i), (ii), (iii) hold if we replace P_{jt} by P^*_{jt} and P_j by P^*_j. Therefore every point P^2 other than P is tamely ramified in K and by Lemma 5, V is nonsingular outside $\phi^{-1}(P)$.

Let $\bar{x} = x/y$ and $\bar{z} = 1/y$. Then $(\bar{x}, \bar{z})$ is a minimal basis of $M(P, P^2)$ and $\bar{x}/\bar{z} = x$. Consequently, $k(x) \subset R_v$ and under the natural epimorphism

[6] $\Delta(B/k[x])$ denotes the discriminant ideal of B over $k[x]$, i.e., the ideal in $k[x]$ generated by the $L/k(x)$-discriminants of all the various $k(x)$-bases of L belonging to B.

$R_v \to R_v/M_v$, $k(x)$ is mapped isomorphically onto R_v/M_v. From this it follows that v has a unique extension w to K, $g(w:v) = [K:k(P^2)]$ and hence w is unramified in K. Since v has center P on P^2, $\phi^{-1}(P)$ consists of a single point. Let L^* be the least galois extension of $k(x)$ in Ω containing L and let $K^* = L^*(y)$. Then K^* is the least galois extension of $k(x,y)$ in Ω containing K and as above, $Q(P,P^2)$ does not split in K^*. Therefore P is tamely ramified in K if and only if $[L^*: k(x)]$ is prime to p, i.e., K/P^2 is tamely ramified if and only if $L \in \Omega^*(P^1 - P_1 - \cdots - P_s)$.

Finally, let u be the map of V^* onto P^2. Since v is unramified in K and since V^* is nonsingular, we have that $\Delta(K/V) \subset u^{-1}[L_1] \cup \cdots \cup u^{-1}[L_s]$. Now $u^{-1}[L_1] \cup \cdots \cup u^{-1}[L_s]$ is nonsingular [A2, Section 7], and hence by Lemma 5, we conclude that K/V^* is tamely ramified.

THEOREM 2. *For L in $\Omega'(P^1 - P_1 - \cdots - P_s)$ let $\tau L = L(y)$, and let $n_1, \cdots, n_s$ be arbitrary nonnegative integers. Then we have the following:* (I) *Given H in $\Omega(P^2 - L_1 - \cdots - L_s)$ such that every point of P^2 other than P is tamely ramified in H, there exists L in $(P^1 - P_1 - \cdots - P_s)$ such that $H = \tau L$.* (II) *For any L in $\Omega'(P^1 - P_1 - \cdots - P_s)$, $[L: k(P^1)] = [\tau L: k(P^2)]$; and $L/k(P^1)$ is galois if and only if $\tau L/k(P^2)$ is galois and then the two galois groups are naturally isomorphic.*

(III) $$\Omega'(P^2 - L_1 - \cdots - L_s) = \Omega^*(P^2 - L_1 - \cdots - L_s)$$

and hence $\bar{\pi}'(P^2 - L_1 - \cdots - L_s) = \bar{\pi}^*(P^2 - L_1 - \cdots - L_s)$.

(IV) *τ maps $\Omega^*(P^1 - P_1 - \cdots - P_s)$ onto $\Omega^*(P^2 - L_1 - \cdots - L_s)$ in a one-to-one manner.*

(V) $$\tau(\Omega^*(P^1 - (P_1, n_1) - \cdots - (P_s, n_s))) = \Omega^*(P^2 - (L_1, n_1) - \cdots - (L_s, n_s)).$$

(VI) *$\bar{\pi}^*(P^1 - P_1 - \cdots - P_s)$ and $\bar{\pi}^*(P^1 - (P_1, n_1) - \cdots - (P_s, n_s))$ are naturally isomorphic respectively to $\bar{\pi}^*(P^2 - L_1 - \cdots - L_s)$* and *$\bar{\pi}^*(P^2 - (L_1, n_1) - \cdots - (L_s, n_s))$.*

Proof.[5] Since $k(x)$ and $k(y)$ are linearly disjoint over k, in view of Proposition 2, it is enough to prove that given H in $\Omega_g(P^2 - L_1 - \cdots - L_s)$ such that every point of P^2 other than P is tamely ramified in H, there exists L in $\Omega'_g(P^1 - P_1 - \cdots - P_s)$ such that $H = L(y)$. Let V be a H-normalization of P^2 and let ϕ be the rational map of V onto P^2. Now

$$\dim |P^1| - \nu(P^1, P^1 \cup L_1 \cup \cdots \cup L_s; P^2) = 2 - 0 > 1,$$

and hence by Lemma 5 and [A2, Proposition 14 of Section 9], $W = \phi^{-1}(P^1)$ is

irreducible [7] and nonsingular. Restricting ϕ to W, W becomes a covering of P^1. Since V/P^2 is tamely ramified except possibly at P, and since $H/k(P^2)$ is galois, the quotient of $[H: k(P^2)]$ by the number of points of V (i.e., of W) lying above any point of P^1 is prime to p. Now $[H: k(P^2)] = [k(W) : k(P^1)]$ and hence the quotient of $[k(W) : k(P^1)]$ by the number of points of W lying above any point of P^1 is prime to p. Since $k(W)/k(P^1)$ is also galois and W is nonsingular, we conclude that W/P^1 is tamely ramified. Also $\Delta(W/P^1) \subset P_1 \cup \cdots \cup P_s$. Let D and E be the residue fields respectively of $Q(P^1, P^2)$ and $Q(W, V)$, and consider D to be a subfield of E in the natural manner. Then E/D is galois and there exist a natural isomorphism σ of D onto $k(x)$, and consequently σ can uniquely be extended to an isomorphism σ^* of E into Ω. Let $L = \sigma^*(E)$. Then $L \in \Omega'_g(P^1 - P_1 - \cdots - P_s)$. Let $H' = L(y)$ and let H^* be the compositum of H' and H^* in Ω. Let V' and V^* be respectively H' and H^* normalizations of P^2, and let ϕ' and ϕ^* be the rational maps respectively of V' and V^* onto P^2. By Proposition 2, $\Delta(H'/P^2) \subset L_1 \cup \cdots \cup L_s$ and every point of P^2 other than P is tamely ramified in H'. Therefore $\Delta(H^*/P^2) \subset L_1 \cup \cdots \cup L_s$ and every point of P^2 other than P is tamely ramified in H^*. Therefore by [A2, Proposition 14 of Section 9] $W^* = \phi^{-1}(P^1)$ and $W' = \phi'^{-1}(P^1)$ are irreducible.[7] Let E^* and E' be the residue fields respectively of $Q(W^*, V^*)$ and $Q(W', V')$. It is clear that E' and E are D-isomorphic. Since P^1 is unramified in H^*, it follows by Lemma 1 that $d(Q(W^*, V^*) : Q(W', V')) = 1$. Since $Q(W^*, V^*)$ is the only local ring in H^* lying above $Q(W', V')$, we conclude that $H^* = H'$.

COROLLARY. *For $j = 1, \cdots, h$ $(h \leqq s)$ let v_j be a valuation of $k(P^2)/k$ of k-dimension one and having center at a point Q_j of L_j other than P and let $n_j = v_j(M(Q_j, L_j, P^2))$. Then*

$$\begin{aligned}
&\bar{\pi}'(P^2 - L_1 - \cdots - L_s + v_1 + \cdots + v_h) \\
&\qquad = \bar{\pi}^*(P^2 - L_1 - \cdots - L_s + v_1 + \cdots + v_h) \\
&\quad = \bar{\pi}'(P^2 - (L_1, n_1) - \cdots - (L_h, n_h) - L_{h+1} - \cdots - L_s) \\
&\qquad = \bar{\pi}^*(P^2 - (L_1, n_1) - \cdots - (L_h, n_h) - L_{h+1} - \cdots - L_s) \\
&\quad = \bar{\pi}^*(P^1 - (P_1, n_1) - \cdots - (P_h, n_h) - P_{h+1} - \cdots - P_s).
\end{aligned}$$

Follows from Theorem 2 and Lemma 5.

---

[7] From the proof of the quoted result it is clear that to deduce the irreducibility of $\phi^{*-1}(P^1)$ (respectively: $\phi^{-1}(P^{-1})$, $\phi'^{-1}(P^1)$), it is enough to know that every point of P^2 which lies on P^1 is tamely ramified in H^* (respectively: H, H').

Remark 3. Let $s \geqq 3$ and assume $p \neq 0, 2$. Then by [A5, Proposition 3 of Section 3], it follows that

$$\Omega^*(P^1 - P_1 - \cdots - P_s) \neq \Omega'(P^1 - P_1 - \cdots - P_s)$$

and in fact that $\bar{\pi}^*(P^1 - P_1 - \cdots - P_s)$ and $\bar{\pi}'(P^1 - P_1 - \cdots - P_s)$ are nonisomorphic. Let V^* be an immediate quadratic transform of P^2 with center at P, and let u be the map of V^* onto P^2. Then from Proposition 2 we deduce that there exist extensions K of $k(P^2)$ such that $\Delta(K/V^*) \subset u^{-1}(L_1 \cup \cdots \cup L_s)$, $\Delta(K/P^2) \subset L_1 \cup \cdots \cup L_s$, K is tamely ramified over V^* but not so over P^2. Note that V^* and P^2 are both nonsingular, $V^* - u^{-1}(L_1 \cup \cdots \cup L_s)$ and $P^2 - L_1 - \cdots - L_s$ are biregularly equivalent, and K is unramified everywhere on these open varieties. *Thus tame ramification is not a birational invariant even for nonsingular models* and more specifically in the present case $V^* - u^{-1}(L_1 \cup \cdots \cup L_s)$ and $P^2 - L_1 - \cdots - L_s$ *are biregularly equivalent but* $\bar{\pi}'(V^* - u^{-1}(L_1 \cup \cdots \cup L_s))$ *and* $\bar{\pi}'(P^2 - L_1 - \cdots - L_s)$ *are not isomorphic,* since by Proposition 2 and Theorem 1 these two groups respectively are isomorphic to

$$\bar{\pi}'(P^1 - P_1 - \cdots - P_s) \text{ and } \bar{\pi}^*(P^1 - P_1 - \cdots - P_s).$$

Another consequence of the above consideration is that P *is the only point of* P^2 *which is non-tamely ramified in* K. *Thus one does not have a theorem of purity for non-tame branch loci on a nonsingular variety.*

THEOREM 3. *We have the following:*

(I) $\Omega'(P^2 - L_1 - \cdots - L_s - P^1) = \Omega^*(P^2 - L_1 - \cdots - L_s - P^1)$

and hence

$$\bar{\pi}'(P^2 - L_1 - \cdots - L_s - P^1) = \bar{\pi}^*(P^2 - L_1 - \cdots - L_s - P^1).$$

(II) $\bar{\Omega}^*(P^2 - L_1 - \cdots - L_s - P^1)$

$=$ *the compositum of* $\Omega^*(P^1 - P_1 - \cdots - P_s)$ *and* $\Omega(k(y), y)$ *in* Ω

$=$ *the compositum of* $\bar{\Omega}^*(P^2 - L_1 - \cdots - L_s)$ *and* $\Omega(k(P^2), y)$ *in* Ω.

(III) $\bar{\Omega}^*(P^2 - L_1 - \cdots - L_s) \cap \Omega(k(P^2), y) = k(P^2)$.

(IV) $\bar{\pi}^*(P^2 - L_1 - \cdots - L_s - P^1) \approx \bar{\pi}^*(P^1 - P_1 - \cdots - P_s) \oplus \bar{Z}^p$.

Proof. Follows from Theorem 2, Proposition 1 and Theorem 1.

THEOREM 4. *Let v_j be a valuation of $k(P^2)/k$ of k-dimension one having center P_j on P^2 such that $v_j(M(P_j,P^1,P^2))=1$ for $j=1,\cdots,h$, with $0<h<s$. Let $n_j=v_j(M(P_j,L_j,P^2))$. Let $m_j=(n_1\cdots n_h)/n_j$. Let d be the greatest common divisor of $m_1,\cdots,m_h$. Let n be the part of $(n_1\cdots n_h)/d$ prime to p. Then*

$$\bar{\pi}'(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_h)$$
$$=\bar{\pi}^*(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_h)$$

and there exists an exact sequence

$$0\to\bar{\pi}^*(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_h)$$
$$\to\bar{\pi}^*(P^1-(P_1,n_1)-\cdots-(P_h,n_h)-P_{h+1}-\cdots-P_s)\oplus\bar{Z}^p$$
$$\to Z_n\to 0.$$

Proof. Relabelling the lines L_j we may assume that L_s is the line at infinity. Then L_j is given by $x=a_j(a_j\in k)$ for $j=1,\cdots,s-1$. Let

$$u=y^{(-n_1\cdots n_h)/d}(x_1-a_1)^{m_1/d}\cdots(x_h-a_h)^{m_h/d}.$$

Then $v_j(u)=0$ for $j=1,\cdots,h$ and for any integer a prime to p we have $\Delta(K(u^{1/a})/P^2)\subset L_1\cup\cdots\cup L_s\cup P^1$, where $K=k(P^2)$. Hence the compositum of $\bar{\Omega}'(P^2-L_1-\cdots-L_s+v_1+\cdots+v_s)$ and $\Omega(K,u)$ is contained in $\bar{\Omega}'(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_s)$. Now let K^* be a member of $\Omega'_g(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_s)$. Let $R=Q(P^1,P^2)$. By [A2, Proposition 14 of Section 9], it follows that there is a unique local ring R^* in K^* lying above R. Let $m=r(R^*:R)$ and let t be an (mn)-th root of u in Ω. Let $L=K(t)$ and $L^*=K^*(t)$. Then L^* belongs to $\Omega'_g(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_s)$. Let S^* be the unique local ring in L^* lying above R. Let K' be the inertia field of S^* over R and let L' be the compositum of K' and L. Then as in the proof of Proposition 1 of Section 3, we deduce that $L'=L^*$. Now we have $K'\in\Omega'_g(P^2-L_1-\cdots-L_s+v_1+\cdots+v_s)$.

Consequently, $\bar{\Omega}'(P^2-L_1-\cdots-L_s-P^1+v_1+\cdots+v_s)$ is the compositum of $\bar{\Omega}'(P^2-L_1-\cdots-L_s+v_1+\cdots+v_s)$ and $\Omega(K,u)$. It is clear that u does not have a q-th root in K for any $q>1$ and that

$$\bar{\Omega}^*(P^2-L_1-\cdots-L_s+v_1+\cdots+v_s)\cap\Omega(K,u)=K(u^{1/n}).$$

Now the theorem follows from Theorems 2 and 3 and Lemma 7.

Remark 4. Let $M_1=P^1$ and let $M_2,\cdots,M_{t-1}$ $(t>0)$ be distinct lines

in P^2 passing through P_1, other than P^1 and L_1. Let S be a line through P_1 other than L_1 and let T be a line through P other than L_1. Let $S_j = S \cap L_j$ and $T_j = T \cap M_j$ and let $T_t = T \cap L_1$. Then $P^2 - L_1 - \cdots - L_s - M_1 - \cdots - M_{t-1}$ can be considered to be the product of $S - S_1 - \cdots - S_s$ and $T - T_1 - \cdots - T_t$. In the classical case, when k is the field of complex numbers, by the product theorem we have

$$\pi_1(P^2 - L_1 - \cdots - L_s - M_1 - \cdots - M_{t-1}) \approx \pi_1(S - S_1 - \cdots - S_s) \oplus \pi_1(T - T_1 - \cdots - T_t).$$

Consequently, in *the abstract case, it can be conjectured that*

$$\bar{\pi}^*(P^2 - L_1 - \cdots - L_s - M_1 - \cdots - M_{t-1}) \approx \bar{\pi}^*(S - S_1 - \cdots - S_s) \oplus \bar{\pi}^*(T - T_1 - \cdots - T_s).$$

Now [A5, Proposition 6 of Section 3], $\bar{\pi}^*(T - T_1) = 1$ and $\bar{\pi}^*(T - T_1 - T_2) = \bar{Z}^p$. Consequently, *the above conjecture has been proved for* $t = 1, 2$ in Theorems 2 and 3 respectively.

However, it follows from Remark 3 ($p \neq 0, 2$; $s \geqq 3$, $t = 1$) *that such a product theorem does not hold for* $\bar{\pi}'$.

Also we note the following. Let

(1) $\bar{\Omega}'(P^2 - L_1 - \cdots - L_s, P)$,

(2) $\bar{\Omega}'(P^2 - (L_1, n_1) - \cdots - (L_h, n_h) - L_{h+1} - \cdots - L_s, P)$,

(3) $\bar{\Omega}'(P^2 - L_1 - \cdots - L_s + v_1 + \cdots + v_s, P)$,

(4) $\bar{\Omega}'(P^2 - L_1 - \cdots - L_s - P^1, P)$,

(5) $\bar{\Omega}'(P^2 - L_1 - \cdots - L_s - P^1 + v_1 + \cdots + v_h, P)$

be the compositums in Ω of all finite algebraic extensions K of $k(P^2)$ such that respectively (1) $\Delta(K/P^2) \subset L_1 \cup \cdots \cup L_s$ and every point of P^2 other than P is tamely ramified in K, (2) $\ldots$, (3) $\ldots$, (4) $\ldots$, (5) $\ldots$, where the notations are as respectively in (1) Theorem 2, (2, 3) Corollary of Theorem 2, (4) Theorem 3, (5) Theorem 4. Let $\bar{\pi}'(P^2 - L_1 - \cdots - L_s, P)$, $\cdots$, denote the corresponding galois groups over $k(P^2)$. Then from the proofs of the quoted results we deduce the following:

(1) $\quad \bar{\pi}'(P^2 - L_1 - \cdots - L_s, P) \approx \bar{\pi}'(P^1 - P_1 - \cdots - P_s)$;

(2, 3) $\quad \bar{\pi}'(P^2 - (L_1, n_1) - \cdots - (L_h, n_h) - L_{h+1} - \cdots - L_s, P)$

$$= \bar{\pi}'(P^2 - L_1 - \cdots - L_s + v_1 + \cdots + v_h, P)$$

$$\approx \bar{\pi}'(P^1 - (P_1, n_1) - \cdots - (P_h, n_h) - P_{h+1} - \cdots - P_s);$$

(4) $\quad \bar{\pi}'(P^2 - L_1 - \cdots - L_s - P^1, P) \approx \bar{\pi}'(P^1 - P_1 - \cdots - P_s) \oplus \bar{Z}^p$;

(5) there exists an exact sequence

$$\begin{aligned}0 &\to \bar{\pi}'(P^2 - L_1 - \cdots - L_s - P^1 + v_1 + \cdots + v_k, P)\\ &\to \bar{\pi}'(P^1 - (P_1, n_1) - \cdots - (P_h, n_h) - P_{h+1} - \cdots - P_s) \oplus Z^p\\ &\to Z_n \to 0.\end{aligned}$$

Remark 5. Let D_1, D_2, D_3 be nonsingular curves in P^2. Let d_j be the order D_j. Let Q be a point of P^2. Assume that $Q = D_1 \cap D_2 = D_1 \cap D_3 = D_2 \cap D_3$ and that D_1, D_2, D_3 are pairwise nontangential at Q. If at least one of the curves D_j is not a line, i. e., if $\max(d_1, d_2, d_3) > 1$, then by Theorem 2 of [A3] it follows that $\bar{\pi}'(P^2 - D_1 - D_2 - D_3)$ is abelian. However, if $d_1 = d_2 = d_3 = 1$, then we can take $s = 3$ and $L_j = D_j$ and then by Theorem 2 it follows that $\bar{\pi}'(P^2 - D_1 - D_2 - D_3) \approx \bar{\pi}^*(P^1 - P_1 - P_2 - P_3)$, and $\bar{\pi}^*(P^1 - P_1 - P_2 - P_3)$ is certainly infinite and unsolvable [A5, Proposition 3 of Section 3] provided $p \neq 2$.

Remark 6. Let P^1 be a projective line over k and let $P_1, \cdots, P_n$ be distinct points in P^1 and let $Q_1, \cdots, Q_n$ be also distinct points in P^1. In view of the situation in the classical case, one expects that

$$\bar{\pi}'(P^1 - P_1 - \cdots - P_n) \approx \bar{\pi}'(P^1 - Q_1 - \cdots - Q_n).$$

This is obviously true for $n \leqq 3$ [A5, Lemma A3 of Section 8].

PRINCETON UNIVERSITY.

REFERENCES.

S. Abhyankar,

[A1] "Tame coverings and fundamental groups of algebraic varieties. Part I: Branch loci with normal crossings; Applications: Theorems of Zariski and Picard," *American Journal of Mathematics*, vol. 81 (1959), pp. 46-94.

[A2] "———. Part II: Branch curves with higher singularities," *ibid.*, vol. 82 (1960), 120-178.

[A3] "———. Part III: Some other sets of conditions for the fundamental group to be abelian," *ibid.*, vol. 82 (1960), pp. 179-190.

[A4] "On the valuations centered in a local domain," *ibid.*, vol. 78 (1956), pp. 321-348.

[A5] "Coverings of algebraic curves," *ibid.*, vol. 79 (1957), pp. 825-856.

[A6] "Local uniformization on algebraic surfaces over ground fields of characteristic $p \neq 0$," *Annals of Mathematics*, vol. 63 (1956), pp. 491-526.

O. Zariski,

[Z] "On the purity of the branch locus of algebraic functions," *Proceedings of the National Academy of Sciences, U. S.A.*, vol. 44 (1958), pp. 791-796.

TAME COVERINGS AND FUNDAMENTAL GROUPS OF ALGEBRAIC VARIETIES.*

Part V: Three Cuspidal Plane Quartics.

By SHREERAM ABHYANKAR.

Introduction. Let H be an irreducible three cuspidal quartic in the projective plane P^2 over an algebraically closed ground field k. For the case when k is the field of complex numbers, using the projection of a cubic surface with a double line and employing some facts from the theory of trisection of periods of elliptic functions, Zariski [Z] found $\pi_1(P^2-H)$ to be a nonabelian group of order 12. In this paper we show that (Theorem 1 and Remark 1) if k is of arbitrary characteristic $p\neq 2,3$ (for $p=2$, H does not exist [A5]), then $\pi'(P^2-H)$ is a nonabelian group of order 12 which is the same as the one found by Zariski. In the abstract case, the theory of trisection of periods of elliptic functions and the topological tools employed by Zariski are not available and consequently our method is entirely geometric. In rough outline it is this. Our first step is the same as that of Zariski, namely we project a cubic surface with a double line to get a three cuspidal quartic as branch locus on a projective plane (all three cuspidal quartics are projectively equivalent [A5]). However, we immediately pass to the corresponding "resolvent surface" (i. e., the associated galois covering) and obtain a birational plane representation of it. Employing the results of [A1, 2] together with a relation between the ramification of a branch curve and the ramification of the P-adic divisor at a simple point P on it [A4, Lemma 5V] and a case of "invariance of tame ramification" [A4, Lemma 6] we show that the "universal covering" of the "resolvent surface" minus the corresponding branch curve is a double covering. The rest is then elementary galois theory. In Remark 2 we examine as to how far we could have gotten by directly using the results of [A1, 2, 3] without the intervention of Lemmas 5V and 6 of [A4]. We remark (Remark 4) that "the cubic surface" and "the resolvent surface" are birationally equivalent since both are rational surfaces.

The preliminary information on cubic surfaces with a double line and the

* Received June 3, 1959.

corresponding "resolvent surface" is contained in [A5] and that paper also contains some applications of the corollary (Theorem 2) of the main result (Theorem 1) of this paper. In [A5] it is also shown that the branch locus of a projection of a cubic surface with a double line need not be a three cuspidal quartic for special positions of the point of projection.

In [Z], Zariski has also shown that $\pi_1(P^2 - H')$ is cyclic for any irreducible curve H' of order at most at most four except when H' is a three cuspidal quartic. The corresponding statement in the abstract case will be proved in Part VI of this series and there we shall also study $\bar{\pi}'(P^2 - H')$ when H' is a reducible curve of order at most four.

We shall use the notations, conventions and definitions of [A1, 2, 3, 4]

1. A lemma.

LEMMA. *Let G be a nonabelian group of order 12 such that there exists an exact sequence $0 \to Z_3 \to G \to Z_4 \to 0$, where Z_4 is a cyclic group of order 4 and Z_3 is a cyclic normal subgroup of G of order 3. Then* (1) *G is generated by two generators α and β having $(\alpha^3 = 1, \beta^4 = 1, \beta\alpha\beta^{-1} = \alpha^2)$ as a complete set of relations. Furthermore,* (2) *G contains unique subgroups G_2, G_4, G_6 of indices 2, 4, 6 respectively*; (3) *G contains exactly three subgroups $G^1{}_3$, $G^2{}_3$, $G^3{}_3$ of index* 3. (4) *G_2, G_4, G_6 are normal in G.* (5) *$G_4 \subset G_2$.* (6) *G/G_2 and G/G_4 are cyclic.* (7) *G/G_6 is isomorphic to the symmetric group on* 3 *letters.* (8) *$G^1{}_3$, $G^2{}_3$, $G^3{}_3$ form a complete set of conjugates in G and they generate G_6.*

Proof. We shall give two proofs of (1). *First proof*: By Hölder's theorem [Zs, Theorem 20 on page 99] G is generated by two generators α and β with the complete set of relations:

$$\alpha^3 = 1, \quad \beta^4 = \alpha^i, \quad \beta\alpha\beta^{-1} = \alpha^j,$$

where i and j are integers such that

$$j^4 \equiv 1 \pmod 3 \text{ and } i(j-1) \equiv 0 \pmod 3.$$

Since $\alpha^3 = 1$ we may take $i = 0$ or 1 or 2 and $j = 0$ or 1 or 2. Since G is nonabelian, we cannot have $j = 1$. Since G is of order greater than 4, we cannot have $j = 0$. Therefore $j = 2$. Hence (mod 3) $0 \equiv i(j-1) \equiv i(2-1) \equiv i$. Therefore $i = 0$. Thus the complete set of relations are: $\alpha^3 = 1$, $\beta^4 = 1$, $\beta\alpha\beta^{-1} = \alpha^2$. *Second proof*: Since 4 is the highest power of 2 which divides 12, G contains a Sylow subgroup Z'_4 of order 4. Since 4 and 3 are coprime,

$Z'_4 \cap Z_3 = 1$ and $Z_3 Z'_4$ is of order divisible by 3 and 4 and hence by 12 and hence $Z_3 Z'_4 = G$. Hence by the homomorphism theorems we have

$$Z'_4 \approx Z'_4/(Z'_4 \cap Z_3) \approx Z_3 Z'_4/Z_3 \approx Z_4.$$

Therefore Z'_4 is cyclic. Let α and β be generators Z_3 and Z'_4, respectively. Then $\alpha^3 = 1$, $\beta^4 = 1$, and $\beta\alpha\beta^{-1} = \alpha^j$ since Z_3 is normal. $j \neq 0$ since $\alpha \neq 1$ and $j \neq 1$ since G is nonabelian. Therefore $j = 2$. The group given by these generators and relations is easily seen to be of order 12 and hence these relations form a complete set.

Since Z_3 is a normal subgroup of G and since 3 is the highest power of 3 which divides 12, Z_3 must be the unique (Sylow) subgroup of G of order 3, i.e., index 4. Thus $G_4 = Z_3$ and $G/Z_3 \approx Z_4$. From this it follows that G_2 is unique, $G_4 \subset G_2$, G_2 is normal in G and G/G_2 is cyclic. By the second proof of (1), it follows that Z'_4 is not normal in G for otherwise G would be a direct product of Z_3 and Z'_4 which would contradict nonabelian nature of G. Let $f\colon G \to S_3$ be the homomorphism of G into S_3 obtained via the left cosets of Z'_4 where S_3 is the symmetric group on three letters. Since Z'_4 is not normal, $f(G)$ cannot be of order 3. Hence f is surjective and the kernel of f is a normal subgroup G_6 of G of index 2 and G_6 is the smallest normal subgroup of G containing Z'_4. Since Z'_4 is a Sylow subgroup of G, we deduce (3) and (8). It now remains to be shown that G_6 is the only subgroup of G of index 2. Let $G^1{}_3 = Z'_4$ and let $G^2{}_3$ and $G^3{}_3$ be the other conjugates of Z'_4. Let G'_6 be any subgroup of G of index 2. By Sylow's theorems, $G'_6 \subset G^j{}_3$ for some j and there exists an inner automorphism g of G such that $g(G^j{}_3) = Z'_4$. Since G_6 is the only subgroup of Z'_4 of order 2, we must have $g(G'_6) = G_6$. Since G_6 is a normal subgroup of G, we must have $G'_6 = G_6$.

2. Results and remarks.

THEOREM 1. *Let P^2 be a projective plane over an algebraically closed ground field k of characteristic $p \neq 3$, let H be an irreducible quartic in P^2 having three cusps, let Ω be an algebraic closure of $K = k(P^2)$. Then* (1) *$\pi'(P^2 - H)$ is a nonabelian group of order* 12 *generated by two generators α and β having $(\alpha^3 = 1,\ \beta^4 = 1,\ \beta\alpha\beta^{-1} = \alpha^2)$ as a complete set of relations;* (2) *$\Omega'(P^2 - H)$ contains unique members K_2, K_4, K_6, K_{12} respectively of degrees* 2, 4, 6, 12 *over K;* (3) *$\Omega'(P^2 - H)$ contains exactly three members $K^1{}_3$, $K^2{}_3$, $K^3{}_3$ of degree* 3 *over K;* (4) *K_2, K_4, K_{12} are galois over K;* (5) *$G(K_2/K)$ and $G(K_4/K)$ are cyclic;* (6) *$G(K_6/K)$ is isomorphic to the*

symmetric group on three letters; (7) $G(K_{12}/K) = \pi'(P^2 - H)$; (8) $K^1{}_3$, $K^2{}_3$, $K^3{}_3$ *are a complete set of K-conjugates in Ω and K_6 is the compositum of* $K^1{}_3$, $K^2{}_3$, $K^3{}_3$ *in* Ω; (9) $K^1{}_3/k$, $K^2{}_3/k$, $K^3{}_3/k$ *and* K_6/k *are all pure transcendental extensions. Also* (10) $p \neq 2$ *and hence* $\pi^*(P^2 - H) = \pi'(P^2 - H)$.

Proof. For $p = 2$ there does not exist any irreducible quartic with three cusps [A5, Proposition 2 of Section 1] and hence we must have $p \neq 2$. Also by [A5, Proposition 2 of Section 1 and Lemma 3 of Section 2] we can choose a k-general affine point (x, z) of P^2 such that H has the affine equation $1 - 4x - 4z - 27x^2z^2 + 18xz = 0$. Let K' be the splitting field over $K = k(P^2) = k(x, z)$ in Ω of the polynomial $T^3 + T^2 + xt + x^2z = 0$. Let V' be a K'-normalization of V and let ϕ be the rational map of V' onto V. Then [A5, Proposition 3 of Section 6] we have the following: (1) $G(K'/K)$ is isomorphic to the symmetric group S_3 on three letters. (2) K'/P^2 is tamely ramified. (3) $\Delta(K'/P^2) = H$. (4) $\phi^{-1}(H)$ has three irreducible components H_1, H_2, H_3. (5) There exists a birational map ψ of V' onto a projective plane P'^2 over k with the following properties: ψ has exactly two fundamental points B and C on V'; $b = \psi(B)$ is an irreducible conic and $c = \psi(C)$ is a line meeting b in two distinct points d_1 and d_2; $h_j = \psi[H_j]$ is a line meeting b in two distinct points a_j and α'_j; h_1, h_2, h_3 have a point γ in common; $a_1, a_2, a_3, \alpha_1, \alpha_2, \alpha_3, \gamma, d_1, d_2$ are all distinct points; a_1, a_2, a_3, d_1, d_2 are the only fundamentl points of ψ^{-1}; $A_1 = \psi^{-1}(a_1)$, $A_2 = \psi^{-1}(a_2)$, $A_3 = \psi^{-1}(a_3)$, $D_1 = \psi^{-1}(d_1)$, $D_2 = \psi^{-1}(d_2)$ are distinct irreducible curves on V' and none of the curves H_j are among these; the real discrete valuation $\mathfrak{a}_j$ of K'/k having center A_j on V' is the a_j-adic divisor of P'^2; and the real discrete valuation $\mathfrak{d}_j$ of K'/k having center D_j on V' is the d_j-adic divisor of P'^2.

Let K_1 be a galois extension of K' in Ω such thot K_1/V' is tamely ramified and $\Delta(K_1/V') \subset H_1 \cup H_2 \cup H_3$. Then

$$\Delta(K_1/P'^2) \subset h_1 \cup h_2 \cup h_3 \cup b \cup c$$

and by [A4, Lemma 6], K_1/P'^2 is tamely ramified. Now

$$\nu(b, h_1 \cup h_2 \cup h_3 \cup b \cup c; P'^2) = 0$$

and $\dim |b| = 5 > 2 = 2 + 0$ and hence [A2, Proposition 14 of Section 9] the real discerte valuation $\mathfrak{b}$ of K'/k having center b on P'^2 has a unique extensions $\mathfrak{b}_1$ to K_1. Let L be the inertia field of $\mathfrak{b}_1$ over $\mathfrak{b}$. Then $\Delta(L/P'^2) \subset h_1 \cup h_2 \cup h_3 \cup c$. Since $A_j \not\subset \Delta(K_1/V')$ and $D_1 \not\subset \Delta(K_1/V')$, $\mathfrak{a}_j$ and $\mathfrak{d}_1$ and unramified in L. Since $\mathfrak{a}_j$ (respectively: $\mathfrak{d}_1$) is the a_j-adic (respectively:

d_1-adic) divisor of P'^2 and since a_j is a simple point of h_j (respectively: c) as well as of $h_1 \cup h_2 \cup h_3 \cup c$, by [A4, Lemma 5V] we conclude that h_j and c are unramified in L and hence L/P'^2 is unramified. Therefore [A1, Section 13] $L = K'$ and hence K_1/K' is cyclic [A1, Lemma 14] of order n, where n is prime to p in case $p \neq 0$. Choose an affine k-general point (u, v) of P'^2/k for which h_1, h_2, h_3, b, c are all at finite distance. Let $\mathfrak{H}_1$, $\mathfrak{H}_2$, $\mathfrak{H}_3$, $\mathfrak{B}$, $\mathfrak{C}$ be irreducible polynomials in $k[u, v]$ which give h_1, h_2, h_3, b, c respectively. Then we can find a primitice element q of K_1/K' such that

$$Q = q^n = \mathfrak{H}_1^{r_1}\mathfrak{H}_2^{r_2}\mathfrak{H}_3^{r_3}\mathfrak{B}^s\mathfrak{C}^t,$$

where r_1, r_2, r_3, s, t are integers [A1, Section 13]. Since the line at infinity on P'^2 is not ramified in K_1, we must have

$$(6) \qquad r_1 + r_2 + r_3 + 2s + t \equiv 0 \pmod{n}.$$

Now $\mathfrak{d}_1(Q) = s + t$, $\mathfrak{a}_j(Q) = r_j + s$. Since $\mathfrak{d}_1$ and $\mathfrak{a}_j$ are unramified in K, we must have

$$(7) \qquad s + t \equiv r_1 + s \equiv r_2 + s \equiv r_3 + s \equiv 0 \pmod{n}$$

i.e., $t \equiv r_1 \equiv r_2 \equiv r_3 \equiv -s \pmod{n}$; and substituting in (6) we obtain, $-2s \equiv 0 \pmod{n}$. Therefore

$$(8) \qquad -2s \equiv 2t \equiv 2r_1 \equiv 2r_2 \equiv 2r_3 \equiv 0 \pmod{n}.$$

Let m be any prime divisor of n. Suppose if possible that $m \neq 2$. Then (8) implies that $s \equiv t \equiv r_1 \equiv r_2 \equiv r_3 \equiv 0 \bmod m)$ and hence $q^{n/m} \in K'$, which is a contradiction since n/m is a positive integer less than n and since $n = [K_1 : K'] = [K'(q) : K']$. Therefore $n = 2^e$, where e is a nonnegative integer. Suppose if possible that $e > 1$. Then (8) implies that $s \equiv t \equiv r_1 \equiv r_2 \equiv r_3 \equiv 0 \pmod{2^{e-1}}$ and hence $q^{n/2} \in K'$ which is a contradiction. Therefore $n = 1$ or $n = 2$.

Now let $Q = \mathfrak{H}_1\mathfrak{H}_2\mathfrak{H}_3\mathfrak{B}^{-1}\mathfrak{C}$, let q be a square root of Q in Ω and take $K_1 = K'(q)$. Then $\Delta(K_1/P'^2) = h_1 \cup h_2 \cup h_3 \cup b \cup c$ because $1 + 1 + 1 - 2 + 1 \equiv 0 \pmod{2}$. Also $\mathfrak{d}_1(Q) = \mathfrak{d}_2(Q) = \mathfrak{a}_1(Q) = \mathfrak{a}_2(Q) = \mathfrak{a}_3(Q) = 0$ and hence $\mathfrak{d}_1$, $\mathfrak{d}_2$, $\mathfrak{a}_1$, $\mathfrak{a}_2$, $\mathfrak{a}_3$ are all unramified in K_1. Therefore by (5), we can conclude that $\Delta(K/V') = H_1 \cup H_2 \cup H_3$.

From these considerations and from the results of Section 2 of [A1] we conclude that K_1 is the compositum of all galois extensions of K in Ω which are tamely ramified over P^2 and for which the branch locus over P^2 is contained in H (and hence coincides with H); and K_1/K is galois and

$G(K_1/K) = \bar{\pi}'(P^2 - H)$ is of order 12. Since $G(K'/K)$ is isomorphic to S_3 which is nonabelian, $\bar{\pi}'(P^2 - H)$ is also nonabelian. Let K'' be the fileld in Ω generated over K by a fourth root of $1 - 4x - 4z - 27x^2z^2 + 18xz$. Then K''/P^2 is tamely ramified and $\Delta(K''/P^2) = H$. Therefore $K'' \subset K_1$. Now K''/K is a galois extension whose galois group is a cyclic group Z_4 of order 4. $G(K_1/K'')$ is a normal subgroup of $G(K_1/K)$ of order 3 and hence $G(K_1/K'')$ is a cyclic group Z_3 of order 3. Thus we have the exact sequence:

$$0 \to Z_3 \to \bar{\pi}'(P^2 - H) \to Z_4 \to 0.$$

The rest now follows from the lemma.

Remark 1. Let the notation be as in Theorem 1. Let $g_1 = \beta$ and $g_3 = \alpha\beta$. Then g_1 and g_3 also generate $\bar{\pi}'(P^2 - H)$. We want to show that (1′) $g_3^2 = g_1^2$, (2′) $g_1^4 = 1$, (3′) $(g_1g_3)^3 = g_1^2$ form a complete set of relations. This will then identity $\bar{\pi}'(P^2 - H)$ with Zariski's calculations [Z, Section 9]. Denote the relations on α β thus: (1) $\alpha^3 = 1$, (2) $\beta^4 = 1$, (3) $\beta\alpha\beta^{-1} = \alpha^2$. *Proof that* $(1,2,3) \Rightarrow (1',2',3')$: Obviously $(2) \Rightarrow (2')$. Also $g_3^2 = \alpha(\alpha^2\beta)\beta$ (by 3) $= \alpha^3\beta^2 = g_1^2$ (by 1). This gives (1′). Now $g_1g_3 = \beta\alpha\beta = \beta\alpha\beta^{-1}\beta^2 = \alpha^2\beta^2$ (by 3); and $(3) \Rightarrow \beta^{-1}\alpha^2\beta = \alpha \Rightarrow \beta^{-1}\alpha\beta = \beta^{-1}\alpha^4\beta = (\beta^{-1}\alpha^2\beta)^2 = \alpha^2 \Rightarrow \beta^{-2}\alpha^2\beta^2$ $= \beta^{-1}(\beta^{-1}\alpha^2\beta)\beta = \beta^{-1}\alpha\beta = \alpha^2 \Rightarrow \alpha^2\beta^2 = \beta^2\alpha^2 \Rightarrow (g_1g_3)^3 = (\alpha^2\beta^2)^3 = \alpha^6\beta^6 = \beta^2$ (by 1 and 2) $= g_1^2$. This gives (3′). *Proof that* $(1',2',3') \Rightarrow (1,2,3)$: Obviously $(2') \Rightarrow (2)$. Also

(4′) $\qquad (1') \Rightarrow \alpha\beta\alpha\beta = \beta^2 \Rightarrow \alpha\beta\alpha\beta^{-1} = 1 \Rightarrow \beta\alpha\beta^{-1} = \alpha^{-1}$.

(2′) and (4′) give

(5′) $$\beta^3\alpha = \alpha^{-1}\beta^3$$

and

(6′) $$\beta\alpha = \alpha^{-1}\beta.$$

Now $(3') \Rightarrow \beta^2 = (\beta\alpha\beta)(\beta\alpha\beta)(\beta\alpha\beta) = \beta\alpha\beta^2\alpha\beta^2\alpha\beta = (\alpha^{-1}\beta)\beta^2\alpha\beta^2\alpha\beta$ (by 6′) $= \alpha^{-1}\beta^3\alpha\beta^2\alpha\beta = \alpha^{-1}(\alpha^{-1}\beta^3)\beta^2\alpha\beta$ (by 5′) $= \alpha^{-2}\beta\alpha\beta$ (by 2′) $= \alpha^{-2}(\alpha^{-1}\beta)\beta$ (by 6′) $= \alpha^{-3}\beta^2 \Rightarrow \alpha^3 = 1$ which gives (1). Also (4′ and 1) $\Rightarrow \beta\alpha\beta^{-1} = \alpha^2$ which gives (3).

Note that, $\bar{\pi}'(P^2 - H)$ found in the above theorem is isomorphic to the group found by Zariski also follows from Theorem 1 directly.

Remark 2. Referring to the proof of Theorem 1, we want to examine how far we can get by directly using the results of [A2] and [A3] without the intervention of [A4, Lemmas 5V and 6]. (1) Since $\bar{\pi}'(P^2 - H)$ turns

out to be nonabelian, Theorem 3 of [A2] could not possibly have been applicable to $P^2 - H$. This is borne out by the inequality $\dim |H| - \nu(H, H; P^2) = 14 - 15 < 2$ [A5, Proposition 2; and A2, Section 7]. (2) By [A5, (48, 49) of Section 6], $\dim |H_j| = 3$ and $\nu(H_j, H_1 \cup H_2 \cup H_3; V') = 3$ and $3 - 3 < 2$ and hence we could not have applied Theorems 1 or 2 of [A2] to $V' - H_1 - H_2 - H_3$. (3) Let K^* be an extension of K obtained by adjoining a root of the polynomial $T^3 + T^2 + xT + x^2z$ in Ω (K' is then the least galois extension of K in Ω containing K^*, see (8), Section 3 of [A5]) let V^* be a K^*-normalization of V and let ϕ^* be the rational map of V^* onto V and let ϕ_1 be the rational map of V' onto V^*. Then [A5, (11, 16, 23, 31, 33, 40) of Section 3 and 6] $\phi^{*-1}(H)$ has two irreducible components H^*_1 and H^*_2 and $\phi_1^{-1}(H^*_1) = H_1$ and $\phi_1^{-1}(H^*_2) = H_2 \cup H_3$; also V^* is simply connected and nonsingular and $\dim |H^*_1| = 5$, $\dim |H^*_2| = 6$, $\nu(H^*_j, H^*_1 \cup H^*_2; V^*) = 6$. Consequently we could not have applied Theorem 1 or Theorem 2 of [A2] to $V^* - H^*_1 - H^*_2$. In view of [A2, Proposition 14 of Section 9], that $\dim |H^*_2| - \nu(H^*_2, H^*_1 \cup H^*_2; V^*) < 2$ is borne out by the fact that H^*_2 splits in K' (and hence in K_1); however H^*_1 does not split in K_1 and consequently we could not have *a priori* predicted that $\dim |H^*_1| - \nu(H^*_1, H^*_1 \cup H^*_2; V^*) < 2$.

Now let us see how much further we get by *allowing the use of* [A4, Lemma 6] *but not that* [A4, Lemma 5V]. (4) Theorem 2 of [A3] is not applicable to $P'^2 - h_1 - h_2 - h_3 - b - c$ since $\dim |h_j| = 2$ and

$$\nu(h_j, h_1 \cup h_2 \cup h_3; P'^2) = 1.$$

(5) By [A: (24), (27), (28), (29), of Section 6 (Correction to (24): B should be replaced by B^*)] there exists a birational map ψ^* of V^* onto a projective plane P^{*2} such that there is only one point B^* of V^* which is fundamental for ψ^*, $b^* = \psi^*(B^*)$ is a line, exactly two points a^*_1 and a^*_2 on P^{*2} are fundamental for ψ^{*-1}, $A^*_1 = \psi^{*-1}(a^*_1)$ and $A^*_2 = \psi^{*-1}(a^*_2)$ are distinct irreducible curves on P^{*2} different from H^*_1 and H^*_2, the valuation α^*_2 of K^*/k having center A^*_2 on V^* is the a^*_2-adic divisor of P^{*2}, $h^*_1 = \psi^*[H^*_1]$ and $h^*_2 = \psi^*[H^*_2]$ are distinct irreducible conics on P^{*2}, h^*_1 and h^*_2 meet at two points neither of which is on b^* and at each of them they have a 2-fold contact, h^*_1 meets b^* in a^*_1 and in another point, h^*_2 meets b^* in a^*_2 and in another point, $\nu(h^*_j, h^*_1 \cup h^*_2; P^{*2}) = 4$ and $\dim |h^*_j| = 5$ and hence $\dim |h^*_j| - \nu(h^*_j, h^*_1 \cup h^*_2; P^{*2}) < 2$. Because of the last inequality we cannot apply Theorem 2 of [A3] to $P^{*2} - h^*_1 - h^*_2 - b^*$.

Finally let us see how far we get by *allowing the use of* [A4, *Lemma* 6]

as well as that of [A4, *Lemma* 5V], *but not passing to the least galois extension* K' (i.e., staying with K^*). (6) Let $\bar{K}$ be a galois extension of K^* in Ω such that $\bar{K}/V^*$ is tamely ramified and $\Delta(\bar{K}/V^*) \subset H^*_1 \cup H^*_2$. Then by [A4, Lemma 6], $\bar{K}/P^{*2}$ is tamely ramified and $\Delta(\bar{K}/P^{*2}) \subset h^*_1 \cup h^*_2 \cup b^*$. Now $\nu(b^*, h^*_1 \cup h^*_2 \cup b^*; P^{*2}) = 0$ and $\dim |b^*| = 2$. Hence by [A2, Proposition 14 of Section 9], the valuation $\mathfrak{b}^*$ of K^*/k having center b^* on P^{*2} has a unique extension $\mathfrak{b}$ to $\bar{K}$. Let $\bar{K}^*$ be the splitting field of $\mathfrak{b}$ over $\mathfrak{b}^*$. Then $\Delta(\bar{K}^*/P^{*2}) \subset h^*_1 \cup h^*_2$. Since a^*_2 is a simple point of h^*_2 as well as $h^*_1 \cup h^*_2$ and $A^*_2 \not\subset H^*_1 \cup H^*_2$, by [A4, Lemma 5V] we conclude that $\Delta(\bar{K}^*/P^{*2}) \subset h^*_1$. Consequently [A1, Section 13], $\bar{K}^*/K^*$ is cyclic of order 1 or 2 and $\bar{K}/\bar{K}^*$ is cyclic. Thus without passing to K' we have found that $\bar{\pi}'(V^* - H^*_1 - H^*_2)$ is either cyclic or metacyclic. (Note that from the proof of Theorem 1 it actually follows that $\bar{\pi}'(V^* - H^*_1 - H^*_2)$ *is cyclic of order* 4.)

THEOREM 2. *Let the notation be as in Theorem* 1. *Let* $\bar{H}$ *be any other irreducible quartic in* P^2 *having three cusps. Let* $\bar{K}^1{}_3$, $\bar{K}^2{}_3$, $\bar{K}^3{}_3$, $\bar{K}_2$, $\bar{K}_4$, $\bar{K}_6$, $\bar{K}_{12}$ *be the respective extension of* K *in* $\Omega'(P^2 - \bar{H})$. *Let* $\bar{V}^1{}_3, \cdots, \bar{V}_{12}$, $V^1{}_3, \cdots, V_{12}$ *be normalizations of* P^2 *in* $\bar{K}^1{}_3, \cdots, K_{12}$, *respectively. Let* $\bar{\phi}^1{}_3, \cdots, \phi_{12}$ *be the rational maps respectively of* $\bar{V}^1{}_3, \cdots, V_{12}$ *onto* P^2. *Then* (1) *there exists a biregular map* (*in fact a projective transformation*) f *of* P^2 *onto itself such that* $f(\bar{H}) = H$. *Now let* f *be given as in* (1). *Then* (2) *there exist biregular maps* $g^1{}_3$, $g^2{}_3$, $g^3{}_3$, g_2, g_4, g_6, g_{12} *respectively of* $\bar{V}^1{}_3, \cdots, \bar{V}_{12}$ *onto respectively* $V^1{}_3, \cdots, V_{12}$ *such that* $\phi^1{}_3 g^1{}_3 = f\bar{\phi}^1{}_3, \cdots, \phi_{12} g_{12} = f\bar{\phi}_{12}$. *Furthermore,* (3) $\bar{K}^1{}_3/k$, $\bar{K}^2{}_3/k$, $\bar{K}^3{}_3/k$ *and* $\bar{K}_6/k$ *are pure transcendental extensions.*

Proof. (1) is proved in Proposition 2 of Section 1 of [A5] and from it (2) and (3) follow by Theorem 1.

Remark 3. Elsewhere we shall show that none of the extensions $\bar{K}_2/k$, $\bar{K}_4/k$, $\bar{K}_{12}/k$ are purely transcendental.

Remark 4. If the assertion in the last paragraph of [Z] which says that ". . . 3 types of *birationally distinct* irreducible algebraic functions, which admit the 3-cuspidal quartic as branch curve, one of order 3, one of order 6 (which is defined by the resolvent of the cubic equation in the previous case), . . ." were to be interpreted as meaning that $\bar{V}^1{}_3$ and $\bar{V}_6$ are birationally distinct, then that would be erroneous, because by Theorem 2, $\bar{V}^1{}_3$ and $\bar{V}_6$ are both rational surfaces (i.e., birationally equivalent to a projective plane).

PRINCETON UNIVERSITY.

REFERENCES.

S. Abhyankar,

[A1] "Tame coverings and fundamental groups of algebraic varieties. Part I: Branch loci with normal crossings; Applications: Theorems of Zariski and Picard," *American Journal of Mathematics*, vol. 81 (1959), pp. 46-94.

[A2] "———. Part II: Branch curves with higher singularities," *ibid.*, vol. 82 (1960), pp. 120-178.

[A3] "———. Part III: Some other sets of condition for the fundamental group to be abelian," *ibid.*, vol. 82 (1960), pp. 179-190.

[A4] "———. Part IV: Product theorems," *ibid.*, vol. 82 (1960), pp. 341-364.

[A5] "Cubic surfaces with a double line," *Memoirs of the College of Science, University of Kyoto, Series A, Mathematics*, vol. 32 (1960), pp. 455-511.

O. Zariski,

[Z] "On the problem of existence of algebraic functions of two variables possessing a given branch curve," *American Journal of Mathematics*, vol. 51 (1929), pp. 305-328.

H. Zassenhaus,

[Zs] The theory of groups, New York, 1949.

TAME COVERINGS AND FUNDAMENTAL GROUPS OF ALGEBRAIC VARIETIES.*

Part VI: Plane Curves of Order at Most Four.

By SHREERAM ABHYANKAR.

Introduction. Throughout the paper we shall use the definitions, results and conventions of [A1, A2, A3, A4]. Also k will denote an algebraically closed ground field of characteristic p, P^2 will denote a projective plane over k, W will denote a curve in P^2, K will denote $k(P^2)$, and Ω will denote an algebraic closure of K.

In this paper we study $\bar{\pi}'(P^2-W)$ under the assumption that the order of W is at most four. The results are summarized here. P^1 here will denote a projective line over k and $P_1, P_2, \cdots,$ will denote distinct points on P^1.

(A). If W is irreducible, then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$ is a cyclic group whose order equals the reduced degree of W; except when W is a three cuspidal quartic.

(1) If W is a three cuspidal quartic, then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$ is a nonabelian group of order 12 provided $p\neq 3$ (this case does not occur for $p=2$).

(B). If W has two irreducible components at least one of which is a line, then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx \bar{Z}^p$ except in the following Case (2).

(2) W consists of a cuspidal cubic together with its flex tangent: Assume $p\neq 2, 3$. Then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$ and there exists an exact sequence

$$0\rightarrow \bar{\pi}^*(P^1-P_1-P_2-P_3)\oplus \bar{Z}^p\rightarrow \bar{\pi}'(P^2-W)\rightarrow S_3\rightarrow 0,$$

where S_3 is the symmetric group on 3 letters. $\bar{\pi}'(P^2-W)$ is unsolvable provided $p\neq 5$.

(C). If W consists of two conics, then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$ and $\bar{\pi}'(P^2-W)\approx Z_2\oplus \bar{Z}^p$ or $\bar{Z}^p$ according as $p\neq 2$ or $p=2$; except in the following case (3).

* Received June 3, 1959.

(3) W consists of two conics either having one common point or having two common tangents:

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\pi^*(P^1-P_1-P_2-(P_3,2)).$$

If $p=2$ then $\bar{\pi}'(P^2-W)\approx\bar{Z}^p$ and if $p\neq 2,3,5$ then $\bar{\pi}'(P^2-W)$ is infinite and unsolvable.

(D). If W has three irreducible components, then

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{Z}^p\oplus\bar{Z}^p$$

except in the following cases (4), (5).

(4) W consists of a conic and two tangents: Then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$; and in case of $p=2$ we have $\bar{\pi}'(P^2-W)=\bar{Z}^p\oplus\bar{Z}^p$; and in case of $p\neq 2$ we have an exact sequence

$$0\rightarrow\bar{\pi}'(P^2-W)\rightarrow\bar{\pi}^*(P^1-P_1-P_2-(P_3,2))\oplus\bar{Z}^p\rightarrow Z_2\rightarrow 0;$$

furthermore $\bar{\pi}'(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.

(5) W consists of three lines through a point:

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-P_3);$$

$\bar{\pi}'(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.

(E). If W consists of four (distinct) lines, then

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{Z}^p\oplus\bar{Z}^p\oplus\bar{Z}^p,$$

provided no three of the lines have a point in common, i.e., except in the following (6), (7).

(6) Three of the lines have a point in common and the fourth line does not pass through it:

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-P_3)\oplus\bar{Z}^p;$$

$\bar{\pi}(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.

(7) All the four lines have a point in common:

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-P_3-P_4);$$

$\bar{\pi}'(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.

The contents of the various sections are clear from their titles.

1. Irreducible curves of order at most four.

LEMMA 1. *Let (R, M) be a complete two-dimensional regular local domain, let A be a one-dimensional prime ideal in R such that $\lambda_R(A) = s > 1$, let a be a generator of A, and let (S, N) be the (unique) immediate quadratic transform of R such that $S^R[A] \neq S$. Then A has a (s-fold) cusp at R if and only if any one of the following two conditions is satisfied:* (1) *If for x in R with $\lambda_R(x) = 1$, we have $A \subset xR + M^{s+1}$, then $A \not\subset xR + M^{s+2}$.* (2) *$\mu_{S,R}(A) = 1$. Furthermore,* (3) *if A does not have a cusp at R, then $2 \leqq \mu_{S,R}(A) \leqq s$ and if $s = 2$, then $S^R[A]$ and MS are nontangential at S.*

Proof. In Section 7 of [A2] it has been proved that condition (1) is sufficient and that condition (2) is necessary. Let a be a generator of A.

To prove the necessity of (1), assume that A has a cusp at R and suppose if possible that for some x in R with $\lambda_R(x) = 1$ we have $a = xu + v$ with $u \in R$ and $v \in M^{s+2}$. By the definition of a cusp, there exists y in R with $\lambda_R(y) = 1$ such that $a \in yR + M^{s+1}$ and $a \notin yR + M^{s+2}$. Now the reduced R-leading form of a equals the s-th power of a linear form times a nonzero constant in R/M. From this, we at once deduce that the reduced R-leading forms of x and y are nonzero constant multiples of this linear form, while the reduced R-leading form of u is a nonzero constant multiple of the $(s-1)$-st power of this linear form. Consequently $u = dx^{s-1} + w$ and $x = ey + q$, where d and e are units in R, $w \in M^s$ and $q \in M^2$. Therefore

$$a = xu + v = x(dx^{s-1} + w) + v = dx^s + xw + v$$
$$= d(ey + q)^s + (ey + q)w + v = ty + dq^s + qw + v, \text{ with } t \in R.$$

Since $q \in M^2$, $s > 1$, $w \in M^s$, $v \in H^{s+2}$, we get $a \in yR + M^{s+2}$ which is a contradiction. This proves the necessity of condition (1). Now the sufficiency of condition (2) follows from (3) and hence it is enough to prove (3). So assume A does not have a cusp at R. Fix x in R with $\lambda_R(x) = 1$ such that $A \subset xR + M^{s+1}$. Then as in the above paragraph we can write

$$a = dx^s + xw + v,\ \lambda_R(d) = 0,\ \lambda_R(w) \geqq s,\ \lambda_R(v) \geqq s + 2.$$

Fix z in R such that (x, z) is a basis of M. Let $x_1 = x/z$. Then (x_1, z) is a basis of N and

$$a = z^s a_1, \text{ where } a_1 = dx_1^s + zw_1x_1 + z^2v_1,$$

where $w_1 = w/z^s$ and $v_1 = v/z^{s+2}$ are in S. Since d is a unit in R, it is a unit also in S. From the above expression, it follows that: (I) a_1 is not divisible by z in S; (II) $2 \leqq \lambda_S(a_1) \leqq s$; and (III) if $s = 2$, then the reduced S-

leading form of a_1 with respect to the basis (x_1, z) of N is not divisible by the reduced S-leading form of z. From these facts (3) follows immediately.

Remark 1. Conditions (1) and (2) in the above lemma give equivalent versions of the definition of a cusp given in Section 7 of [A2]. Also the definition given in Section 7 of [A2] for a cusp of a curve in a projective plane to be ordinary for that plane is superfluous because it follows from the above lemma that this is always the case.

LEMMA 2. *Assume that W is irreducible of order $w \leqq 4$. Then* $\dim |W| > 1 + \nu(W, W; P^2)$ *except in the following two cases. Case* (I) : $w = 4$ *and W has exactly two singularities; at one of them W has a 2-fold cusp and at the other W is analytically irreducible and has a double point which is not a cusp; we have* $\dim |W| = 14$ *and* $\nu(W, W; P^2) = 13$. *Case* (II) : *W has exactly three singularities and at each of them it has a 2-fold cusp; we have* $\dim |W| = 14$ *and* $\nu(W, W; P^2) = 15$.

Proof. For $w < 4$, the statement has been proved in [A2, Example 3 of Section 10]. Now assume that $w = 4$. Then $\dim W = 14$ and hence we have to show that except in cases (I) and (II), we have $\nu(W, W; P^2) \leqq 12$. Now W has at most three singularities and if W has a triple point, then it is the only singularity of W. Consequently, we have the following possibilities: *Case* 1: W has only one singular point P and it is a triple point. *Case* 2: W has only one singular point P and it is a double point. *Case* 3: W has only two singular points P_1 and P_2 and they are double points. *Case* 4: W has only three singular points P_1, P_2, P_3 and they are double points. We shall deal with these cases separately.

Case 1. Now $\nu(W, W; P^2) = \nu(W, W; P, P^2)$. We shall show that $\nu(W, W; P, P^2) \leqq 9$. Let $W_1, \cdots, W_t$ $(t \leqq 3)$ be the distinct analytic branches of W at P. We have to consider three subcases according as $t = 1, 2, 3$ respectively. *Case* 1A: $t = 1$. Then there is a unique tangent line L to W at P and $i(L \cdot W_1; P, P^2) = i(L \cdot W; P, P^2) = 4$ and hence W has a 3-fold cusp at P. Therefore $\nu(W, W; P, P^2) = (\frac{1}{2})(3)(3+3) = 9$. *Case* 1B: $t = 2$. We can label W_1, W_2 so that W_j has a j-fold point at P. Let L be the tangent line to W_2 at P. Then $4 \leqq i(L \cdot W; P, P^2) = i(L \cdot W_1; P, P^2) + i(L \cdot W_2; P, P^2)$ and $i(L \cdot W_1; P, P^2) \geqq 1$ and $i(L \cdot W_2; P, P^2) \geqq 3$. Therefore $i(L \cdot W_1; P, P^2) = 1$ and $i(L \cdot W_2; P, P^2) = 3$, i.e., W_2 has a 2-fold cusp at P, and W_1 and W_2 are nontangential at P. Therefore $\nu(W, W; P, P^2) = (\frac{1}{2})(3)(3+1) + 2 = 8$. *Case* 1C: $t = 3$. As in Case 1B we can deduce that W_1, W_2, W_3 are pairwise nontangential at P, i.e., W has an ordinary 3-fold point at P and hence $\nu(W, W; P, P^2) = (\frac{1}{2})(3)(3+1) = 6$.

Case 2. Again $\nu(W,W;P^2)=\nu(W,W;P,P)$. We shall show that $\nu(W,W;P,P^2)\leqq 11$. We have the following three subcases. *Case* 2A: *W has two analytic branches* W_1 *and* W_2 *at* P. Let the contact of W_1 and W_2 at P be s-fold. Then using the results in Section 7 of [A2], it can at once be seen that besides P, W has $s-1$ double points infinitely near P. Consequently $s\leqq 3$. [If an irreducible plane curve of order n has singularities of multiplicities $n_1,n_2,\cdots$ at ordinary as well as infinitely near points, then $(n-1)(n-2)\geqq\sum n_i(n_i-1)$. The classical proof of this inequality applies to algebraically closed ground fields of arbitrary characteristic.] Hence $\nu(W,W;P,P^2)\leqq 9$. *Case* 2B: W is *analytically irreducible at* P *and has a cusp* (2-fold) *at* P. Then $\nu(W,W;P,P^2)=5$. *Case* 2C: W *is analytically irreducible at* P *but does not have a cusp there.* Let (R,M) be the quotient ring of P on P^2 and let (R_j,M_j) be the (unique) j-th quadratic transform of R such that $A_j=R_j{}^R[A]\neq R_j$, where A is the ideal of W at P. By Lemma 1, $\lambda_{R_1}(A_1)=2$ and A_1 and MR_1 are nontangential at R_1. If A_1 has a cusp at R_1, then

$$\begin{aligned}\nu(W,W;P,P^2)=\nu(A,A;R)&=\nu(A,A;R,R)+\nu(A_1,A_1;S)\\&=3+5=8.\end{aligned}$$

Now assume that A_1 does not have a cusp at R_1. Then by Lemma 1, $\lambda_{R_2}(A_2)=2$ and A_2 and M_1R_2 are nontangential at R_2 and we also have that $R_2{}^{R_1}[MR_1]=R_2$. If A_2 did not have a cusp at R_2, then by Lemma 1, $\lambda_{R_3}(A_3)$ would equal 2 and hence W would have a double point at P and double points at least at three points infinitely near to P and this would contradict the irreducibility of W. Therefore A_2 must have a cusp at R_2 and hence

$$\begin{aligned}\nu(W,W;P,P^2)&=\nu(A,A;R,R)+\nu(A,A;R_1,R)+\nu(A_2,A_2;R_2)\\&=3+3+5=11.\end{aligned}$$

Case 3. We have the following six subcases. In all subcases except in Case 3F (which is Case I) we shall show that $\nu(W,W;P^2)\leqq 11$.

Case 3A: W *is analytically reducible at* P_1 *and* P_2, *and has a normal crossing at least at one of them.* By suitably labelling P_1 and P_2 assume that W has a normal crossing at P_1 and let s be the order of contact of the branches of W at P_2. Suppose if possible that $s>2$. Then besides at P_1 and P_2, W has multiple points at least at two points infinitely near P_2. This contradicts the irreducibility of W. Therefore $s\leqq 2$ and hence

$$\nu(W,W;P^2)=\nu(W,W;P_1,P^2)+\nu(W,W;P_2,P^2)\leqq 3+6=9.$$

Case 3B: *W is analytically reducible at P_1 and P_2 and does not have a normal crossing at either of them.* Then besides at P_1 and P_2, W has a double point at least at one point infinitely near to P_j for $j=1,2$. This contradicts the irreducibility of W.

Case 3C: *W has a normal crossing at one of the points P_1, P_2 and is analytically irreducible at the other.* Suitably labelling P_1, P_2, we may assume that W has a normal crossing at P_1. Let P_3 be the point in the first neighborhood of P_2 through which the transform of W passes. If P_3 were neither a simple point nor a cusp of W, then by Lemma 1 there would be a singular point of W infinitely near P_3 and this would contradict the irreducibility of W. Therefore

$$\nu(W,W;P^2)\leqq 3+3+5=11.$$

Case 3D: *W is analytically reducible at one of the points P_1, P_2 but does not have a normal crossing there and is analytically irreducible at the other.* Suitably labelling P_1, P_2, we may assume that W is analytically reducible at P_1. Since W is irreducible, the two branches at P_1 must have a 2-fold contact and by Lemma 1, W must have a cusp at P_2. Hence

$$\nu(W,W;P^2)=6+5=11.$$

Case 3E: *W has a cusp at P_1 and P_2.* Then $\nu(W,W;P^2)=10$.

Case 3F: (*i.e., Case* I): *W has a cusp at one of the points P_1, P_2 and is analytically irreducible at the other but does not have a cusp there.* Suitably labelling P_1, P_2, we may assume that W has a cusp at P_1. Let P_3 be the point in the first neighborhood of P_2 through which the transform of W passes. Since W is irreducible, by Lemma 1, W must have a 2-fold cusp at P_3 and hence $\nu(W,W;P^2)=5+3+5=13$.

Case 4: By Proposition 2 of [A5], either we have Case (II) or W has a normal crossing at each of the points P_1, P_2, P_3 so that $\nu(W,W;P^2)=9$.

THEOREM 1. *Assume that W is irreducible of order $w\leqq 4$ and let w^* be the reduced order of W, i.e., $w^*=w$ in case $p=0$ and $w^*=$ the part of w prime to p in case $p\neq 0$. Then $\bar{\pi}'(P^2-W)$ is cyclic of order w^* provided W is not a three cuspidal quartic. If W is a three cuspidal quartic and $p\neq 3$, then $\bar{\pi}'(P^2-W)$ is the (unique) nonabelian extension of a cyclic group Z_3 of order 3 by a cyclic group Z_4 of order 4, i.e., there is an exact sequence $0\to Z_3\to\bar{\pi}'(P^2-W)\to Z_4\to 0$. Furthermore, we always have $\bar{\pi}^*(P^2-W)=\bar{\pi}'(P^2-W)$.*

Proof. In the notation of Lemma 2 suppose we have neither Case I nor Case II; then by Lemma 2 our assertion follows by Theorem 3 of [A2]; also Case II has been dealt with in Theorem 1 of [A5]. Thus we may now assume that W is an irreducible quartic having exactly two singularities P and Q and that at P, W has a 2-fold cusp and at Q, W is analytically irreducible and has a double point which is not a cusp; and furthermore that $\nu(W,W;P^2)=13$. Now $\dim|W|=14 \not> 13+1$ and hence Theorem 3 of [A2] is not directly applicable.

Now the tangent to W at any one of the points P, Q cannot pass through the other point and hence we can find projective coordinates X, Y, Z in P^2 such that P is $X=Z=0$, the tangent to W at P is $Z=0$, Q is $X=Y=0$, and the tangent to W at Q is $Y=0$. Then W is given by $f=0$, where

$$f(X,Y,Z)=Y^2Z^2+aX^2YZ+bX^4+cX^3Y;\ a,b,c\in k;\ b\neq 0\neq c.$$

Let $\bar{X}$, $\bar{Y}$, $\bar{Z}$ be projective coordinates in another projective plane $\bar{P}^2$ over k and let τ be the birational transformation of P^2 onto $\bar{P}^2$ given by [1]

$$X:Y:Z=\bar{X}\bar{Y}:\bar{X}^2:\bar{Y}\bar{Z};\qquad \bar{X}:\bar{Y}:\bar{Z}=XY:X^2:YZ.$$

Now $f(X,Y,Z)=\bar{f}(\bar{X},\bar{Y},\bar{Z})=\bar{Z}^2+a\bar{Y}\bar{Z}+b\bar{Y}^2+c\bar{X}\bar{Y}$. $\bar{f}$ is divisible neither by $\bar{X}$ nor by $\bar{Y}$ and hence $\bar{W}=\tau[W]$ is given by $\bar{f}=0$. The fundamental points of τ are P and Q, $\tau(P)=(\bar{Y}=0)$, $\tau(Q)=(\bar{X}=0)$. Also τ^{-1} is regular everywhere except at $\bar{P}:(\bar{X}=\bar{Z}=0)$ and $\bar{Q}:(\bar{X}=\bar{Y}=0)$, the valuation of K/k having center at $(Y=0)$ on P^2 is the $\bar{P}$-adic divisor of $\bar{P}^2$, and $(Y=0)\not\subset W$. Furthermore, $\bar{P}\notin\bar{W}$ and $\bar{Q}\notin\bar{W}$, and hence $\bar{W}\cup(\bar{X}=0)\cup(\bar{Y}=0)$ has a strong normal crossing at these points. Hence by Lemma 6 of [A4],

$$\Omega'_g(P^2-W)\subset\Omega'_g(\bar{P}^2-\bar{W}-(\bar{X}=0)-(\bar{Y}=0)+(\text{the }\bar{P}\text{-adic divisor of }\bar{P}^2)).$$

Now $\bar{W}\cup(\bar{X}=0)\cup(\bar{Y}=0)$, as well as $(\bar{X}=0)$, has a simple point at $\bar{P}$ and hence by Lemma 5 of [A4], the right-hand side of the above inclusion equals $\Omega'_g(\bar{P}^2-\bar{W}-(\bar{Y}=0))$. Therefore

$$\Omega'_g(P^2-W)\subset\Omega'_g(\bar{P}^2-\bar{W}-(\bar{Y}=0)).$$

Next, $\bar{W}$ is nonsingular and has a 2-fold contact with $\bar{Y}=0$ at $\bar{Y}=\bar{Z}=0$ and hence

---

[1] τ is a Cremona quadratic transformation of the second kind given by the net of conics passing through P and Q and having $Y=0$ as tangent at Q. For the properties of τ to be used here, see Lemma 9 of Section 5 of [A6].

$$\dim|\bar{W}| - \nu(\bar{W}, \bar{W} \cup (\bar{Y}=0); \bar{P}^2) = 5 - 2 > 1; \text{ and}$$
$$\dim|(\bar{Y}=0)| - \nu((\bar{Y}=0), (\bar{Y}=0); \bar{P}^2) = 2 > 1.$$

Therefore by Theorem 2 of [A3], $\pi'(\bar{P}^2 - \bar{W} - (\bar{Y}=0))$ is abelian and hence by the inclusion, $\pi'(P^2 - W)$ is also abelian. Consequently by Lemma 36 of Section 14 of [A1], $\pi'(P^2 - W)$ is cyclic of order w^*.

2. A cubic and a line.

LEMMA 3. *Assume that W consists of an irreducible cubic H and a line A. If H is a cuspidal cubic and A is a flex tangent of H, then $\nu(H, W; P^2) = 8$ and $\nu(A, W; P^2) = 3$, while in all other cases we have $\nu(H, W; P^2) \leqq 7$.*

Proof. We have the following possibilities: *Case* 1, H is nonsingular; *Case* 2, H has an ordinary double point P; *Case* 3, H has a 2-fold cusp at a point P and A is not a flex tangent of H; *Case* 4, H has a 2-fold cusp at a point P and A is a flex tangent of H at a point Q. Note that in Cases 2, 3, 4, P is the only singular point of H.

Case 1. If A meets H in three distinct points, then A is nowhere tangential to H and hence $\nu(H, W; P^2) = 0$. If A meets H in two distinct points P_1 and P_2, then after a suitable labelling of P_j we have that H and A have a j-fold contact at P_j and hence $\nu(H, W; P^2) = 2$. If A meets H in a single point P, then at P, A and H have a three fold contact and $\nu(H, W; P^2) = 3$.

Case 2. If A is not tangent to H at P, then as in Case 1 we get $\nu(H, W; P^2) \leqq 3 + \nu(H, H; P^2) = 3 + 3 = 6$. Now assume that A is tangent to H at P. Then A has a 2-fold contact with one branch of H at P and is nontangential to the other. Hence $\nu(H, W; P^2) = \nu(H, W; P, P^2) = 3 + 1 = 4$.

Case 3. By the results of [A2, Section 7] it follows that $\nu(H, W; P, P^2) = 5$ (even in case A passes through P and even in case A passes through P and is tangent to H at P). Hence $\nu(H, W; P, P^2) \leqq 5 + 2 = 7$.

Case 4. It is clear that in this case $A \cap H = Q$ and that A and H have a 3-fold contact at Q. Hence $\nu(H, W; P^2) = 5 + 3 = 8$ and $\nu(A, W; P^2) = 3$.

THEOREM 2. *Assume that W consists of an irreducible cubic H and a line A. Unless H is a cuspidal cubic and A is its flex tangent, we have $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W) \approx \bar{Z}^p$. If H is a cuspidal cubic and A is its flex*

tangent and if $p \neq 2, 3$, *then* $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W)$ *and there exists an exact sequence*

$$0 \to \bar{\pi}^*(P^1 - P_1 - P_2 - P_3) \oplus Z_p \to \bar{\pi}'(P^2 - W) \to S_3 \to 0,$$

where P_1, P_2, P_3 *are distinct points on a projective line* P^1 *over* k, *and* S_3 *is the symmetric group on 3 letters; in particular,* $\bar{\pi}'(P^2 - W)$ *is infinite and unsolvable provided* $p \neq 2, 3, 5$.

Proof. In the first case the assertion follows from Lemma 3 and [A3, Theorem 2]. Now assume that H is a cuspidal cubic and A is its flex tangent and that $p \neq 2, 3$. We can choose affine coordinates x, z in P^2 such that H and A are given respectively by $4 + 27z^2x = 0$ and $x = 0$ [A6, Lemma 1 of Section 1]. Let K^* be a splitting field over $K = k(P^2) = k(x, z)$ of the polynomial $T^3 + xT + x^2z = 0$. Let V be a K^*-normalization of P^2 and let ϕ be the map of V onto P^2. Then [A6, Proposition 4 of Section 7] $K^* \in \Omega^*(P^2 - W)$, $G(K^*/K) \approx S_3$, $\phi^{-1}[H]$ has three irreducible components H_1, H_2, H_3; $\phi^{-1}[A]$ has three irreducible components A_1, A_2, A_3; $H_1 \cap H_2 \cap H_3 \cap A_1 \cap A_2 \cap A_3$ is a point B which is the only singularity of V, and there exists a birational map ψ of V onto a projective plane P^{*2} with the following properties: (1) ψ is biregular on $V - A_1 - A_2 - A_3$ and maps it onto $P^{*2} - b$, where b is a line on P^{*2}; (2) B is the only fundamental point of ψ and $\psi(B) = b$; (3) ψ^{-1} has exactly three fundamental points a_1, a_2, a_3 and they all lie on b, and after a suitable labelling of them we have $\psi^{-1}(a_j) = A_j$; (4) $h_1 = \psi[H_1]$, $h_2 = \psi[H_2]$, $h_3 = \psi[H_3]$ are distinct lines in P^{*2} other than b; (5) h_1, h_2, h_3 have a point d in common; (6) the points a_1, a_2, a_3, d, $h_1 \cap b$, $h_2 \cap b$, $h_3 \cap b$ are all distinct.

It is clear that

$$\Omega_g(V - H_1 - H_2 - H_3 - A_1 - A_2 - A_3) = \Omega_g(P^{*2} - h_1 - h_2 - h_3 - b).$$

Let K_1 be a member of $\Omega'_g(V - H_1 - H_2 - H_3 - A_1 - A_2 - A_3)$. Then by [A4, Lemma 6 of Section 2], K_1/P^{*2} is tamely ramified. Consequently, by [A2, Proposition 14 of Section 9], the valuation of K^*/k having center b on P^{*2} has a unique extension in K_1. This valuation has center B on V and hence there exists a unique local ring R_1 in K_1 lying above $R = Q(B, V)$. Therefore $G_i(R_1 \colon R) = G(K_1/K^*)$. Since B is tamely ramified in K_1, we conclude that $[K_1 \colon K^*]$ is prime to p. Alternatively, since

$$K_1 \in \Omega'_g(P^{*2} - h_1 - h_2 - h_3 - b),$$

by [A4, Theorem 3 of Section 4], it follows that $[K_1 \colon K^*]$ is prime to p. It is obvious that

$$\Omega^*_g(P^{*2} - h_1 - h_2 - h_3 - b) = \Omega^*_g(V - H_1 - H_2 - H_3 - A_1 - A_2 - A_3)$$

and hence we conclude that

$$\begin{aligned}\bar{\Omega}'(V-H_1-H_2-H_3-A_1-A_2-A_3)\\ &=\bar{\Omega}^*(V-H_1-H_2-H_3-A_1-A\ -A_3)\\ &=\bar{\Omega}^*(P^2-h_1-h_2-h_3-b).\end{aligned}$$

The required exact sequence now follows from [A4, Theorem 3 of Section 4]. Also, by [A7, Proposition 3 of Section 3] it follows that $\bar{\pi}^*(P^1-P_1-P_2-P_3)$ and hence $\bar{\pi}'(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.

3. Two conics.

THEOREM 3. *Assume that W consists of two distinct irreducible conics H_1 and H_2. Then we have the following possibilities:* (1) *H_1 and H_2 meet in four distinct points*; (2) *H_1 and H_2 meet in three distinct points*; (3) *H_1 and H_2 meet in two distinct points and are tangential at only one of them*; (4) *H_1 and H_2 meet in two distinct points and are tangential at both of them*; (5) *H_1 and H_2 meet in only one point. In cases* (1), (2), *and* (3) *we have $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$ and $\bar{\pi}'(P^2-W)\approx Z_2\oplus\bar{Z}^p$ if $p\neq 2$ and $\bar{\pi}'(P^2-W)\approx\bar{Z}^p$ if $p=2$. In cases* (4) *and* (5) *we have $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-(P_3,2))$, where P^1 is a projective line over k and P_1,P_2,P_3 are distinct points on P^1; consequently $\bar{\pi}'(P^2-W)\approx\bar{Z}^p$ in case $p=2$ and $\bar{\pi}'(P-W)$ is infinite and unsolvable in case $p\neq 2,3,5$.*

Proof. In cases $(1,2,3)$, we have $\nu(H_j,W;P^2)\leqq 3$ and $\dim|H_j|=5$ and hence the results follow from [A2, Theorem 3 of Section 10] or [A4, Corollary 4 of Theorem 1 of Section 3]. In cases (4) and (5), we have $\nu(H_j,W;P^2)=4$ and hence these results do not apply, and we proceed as follows.

Case 4. Choose projective coordinates X, Y, Z in P^2 such that the common points of H_1 and H_2 are P_1: $X=Y=0$ and P_2: $X=Z=0$ and the tangents at these points are $Y=0$ and $Z=0$ respectively. Then Q_j is given by $X^2+a_jYZ=0$, where a_1 and a_2 are distinct nonzero elements in k. Let $\bar{X}$, $\bar{Y}$, $\bar{Z}$ be projective coordinates in another projective plane $\bar{P}^2$ over k and let τ be the birational transformation of P^2 onto $\bar{P}^2$ given by [1]:

$$X:Y:Z=\bar{X}\bar{Y}:\bar{X}^2:\bar{Y}\bar{Z};\qquad \bar{X}:\bar{Y}:\bar{Z}=XY:X^2:YZ.$$

Then $\tau[Q_j]$ is the line $\bar{H}_j$ given by $\bar{Y}+a_j\bar{Z}=0$; the only fundamental points of τ are P_1 and P_2 and $\tau(P_j)=\bar{A}_j$, where $\bar{A}_1$ is the line $\bar{X}=0$ and $\bar{A}_2$ is the line $\bar{Y}=0$; the only fundamental points of τ^{-1} are $\bar{P}_1$: $\bar{X}=\bar{Y}=0$ and

$\bar{P}_2\colon \bar{X}=\bar{Z}=0$, and $\tau^{-1}(P_j)=A_j$, where A_1 is the line $X=0$ and A_2 is the line $Y=0$. Let v_j be the valuation of K/k having center A_j on P^2; then v_2 is the $\bar{P}_2$-adic divisor of $\bar{P}^2$.

Let K^* be a member of $\Omega'_g(P^2-W)$. Then

$$\Delta(K^*/\bar{P}^2)\subset \bar{H}_1\cup\bar{H}_2\cup\bar{A}_1\cup\bar{A}_2.$$

Now $\bar{H}_1\cup\bar{H}_2\cup\bar{A}_1\cup\bar{A}_2$ has a strong normal crossing at every point of $\bar{A}_1$ and both the fundamental points P_1 and $\bar{P}_2$ of τ^{-1} are on $\bar{A}_1$. Hence by [A4, Lemma 6 of Section 2], $K^*/\bar{P}^2$ is tamely ramified and hence by [A2, Proposition 14 of Section 9], the valuation of K/k having center $\bar{A}_1$ on $\bar{P}^2$ has a unique extension to K^*. This valuation has center P_1 on P^2 and hence there is a unique local ring R^* in K^* lying above $R=Q(P_1;P^2)$ and hence $G_i(R^*/R)=G(K^*/K)$. Since R is tamely ramified in K^*, we conclude that $[K^*:K]$ is prime to p. Thus

$$\bar{\Omega}'(P^2-W)=\bar{\Omega}^*(P^2-W)\subset\bar{\Omega}^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-\bar{A}_1-\bar{A}_2).$$

Consequently

$$\bar{\Omega}'(P^2-W)=\bar{\Omega}^*(P^2-W)=\bar{\Omega}^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-\bar{A}_1-\bar{A}_2+v_1+v_2).$$

Since v_2 is the $\bar{P}_2$-adic divisor of $\bar{P}^2$ and $\bar{P}_2$ is a simple point of $\bar{H}_1\cup\bar{H}_2\cup\bar{A}_1\cup\bar{A}_2$, as well as of $\bar{A}_1$, by [A4, Lemma 5 of Section 2] we can conclude that

$$\bar{\Omega}^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-\bar{A}_1-\bar{A}_2+v_1+v_2)=\bar{\Omega}^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-\bar{A}_2+v_1).$$

Let $x=X/Z$, $y=Y/Z$, $\bar{x}=\bar{X}/\bar{Z}$, $\bar{y}=\bar{Y}/\bar{Z}$. Then $\bar{x}=x$, $\bar{y}=x^2/y$, $v_1(x)=1$, $v_1(y)=0$. Hence $v_1(\bar{y})=2$. By [A4, Lemma 5 of Section 2] we now deduce that $\Omega^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-\bar{A}_2+v_1)=\Omega^*(\bar{P}^2-\bar{H}_1-\bar{H}_2-(\bar{A}_2,2))$. Hence by [A4, Theorem 2 of Section 4] we conclude that

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-(P_3,2)),$$

where P^1 is a projective line and P_1, P_2, P_3 are distinct points on P^1.

Case 5. Let $P=Q_1\cap Q_2$, let A be the common tangent of Q_1 and Q_2 at P, let P^* be another point of Q_1 and let A^* be the tangent to Q_1 at P^*. Choose projective coordinates X, Y, Z in P^2 such that $P\colon X=Z=0$, $P^*\colon Y=Z=0$, $A\colon X=0$, $A^*\colon Y=0$. Then Q_1 is given by $Z^2+aXY=0$, $0\neq a\in k$. Via a change of coordinates replacing aY by Y, Q_1 is given by $f_1=0$, where $f_1(X,Y,Z)=Z^2+XY$. Now Q_2 is given by $f_2=0$, where $f_2(X,Y,Z)=bZ^2+cZX+dX^2+eXY$. Since $i(Q_1,Q_2;P,P^2)=4$, $f_2(-t^2,1,t)=bt^2-ct^3+dt^4-et^2$ must have $t=0$ as a root of multiplicity

four. Hence $b-e=c=0$ and $d\neq 0$. Dividing f_2 by b we can thus take $f_2(X,Y,Z)=Z^2+XY+uX^2$, $0\neq u\in k$. Let $\bar{P}^2$ be a projective plane with coordinates $\bar{X}$, $\bar{Y}$, $\bar{Z}$. Let $x=X/Z$, $y=Y/Z$, $\bar{x}=\bar{X}/\bar{Z}$, $\bar{y}=\bar{Y}/\bar{Z}$. Let τ be the birational map of P^2 onto $\bar{P}^2$ obtained by setting

(1) $\qquad \bar{x}=x^{-1}+y,\ \bar{y}=x.$ Then $x=\bar{y},\ y=\bar{x}-\bar{y}^{-1}$.

In homogeneous coordinates we have [2]:

[2] In geometric language, τ is a Cremona quadratic transformation of the third kind given by the net of conics osculating at P with Q_1, i. e., having at least a 3-fold contact with Q_1 at P.

(2) $\quad \bar{X}:\bar{Y}:\bar{Z}=Z^2+XY:X^2:XZ; \qquad X:Y:Z=\bar{Y}^2:\bar{X}\bar{Y}-\bar{Z}^2:\bar{Y}\bar{Z}.$

Let P be the point: $\bar{Y}=\bar{Z}=0$ and let $\bar{A}$ be the line: $\bar{Y}=0$. Let $y_1=y/x$, $z_1=1/x$. Let $\bar{x}_1=\bar{x}/\bar{y}$, $\bar{z}_1=1/\bar{y}$. Then

(3) $\qquad \bar{x}_1=z_1^2+y_1,\ \bar{z}_1=z_1$; and $y_1=\bar{x}_1-\bar{z}_1^2,\ z_1=\bar{z}_1$.

From equations (3), it follows that τ is biregular on P^2-A and maps it onto $\bar{P}^2-\bar{A}$. Let v and $\bar{v}$ be the valuations of $k(P^2)/k=k(\bar{P}^2)/k$ having centers A and $\bar{A}$ on P^2 and $\bar{P}^2$ respectively. Then $v(x)=1$ and $v(y)=0$, and hence $v(\bar{y}/\bar{x})>0$ and $v(1/\bar{x})>0$. Hence v has center $\bar{P}$ on $\bar{P}^2$. Also $\bar{v}(\bar{y})=1$ and $\bar{v}(\bar{x})=0$, and hence $\bar{v}(x/y)>0$ and $\bar{v}(1/y)>0$. Hence $\bar{v}$ has center P on P^2. From this we deduce that P is the only fundamental point of τ, $\tau(P)=\bar{A}$, $\tau[A]=\bar{P}$; and $\bar{P}$ is the only fundamental point of τ^{-1}, $\tau^{-1}(\bar{P})=A$, $\tau^{-1}[\bar{A}]=P$.

Now

$$f_1(X,Y,Z)=\bar{X} \text{ and } f_2(X,Y,Z)=\bar{X}+u\bar{Y}.$$

Hence $\bar{Q}_1=\tau[Q_1]$ and $\bar{Q}_2=\tau[Q_2]$ are the lines $\bar{X}=0$ and $\bar{X}+u\bar{Y}=0$ respectively. Now the lines $\bar{Q}_1$, $\bar{Q}_2$, $\bar{A}$ have the common point $\bar{X}=\bar{Y}=0$ which is different from $\bar{P}$. Hence by [A4, Lemma 6 of Section 2] we deduce that

$$\bar{\Omega}'(P^2-W)\subset\bar{\Omega}'(\bar{P}^2-\bar{Q}_1-\bar{Q}_2-\bar{A}).$$

By [A4, Theorem 2 of Section 4],

$$\bar{\Omega}'(\bar{P}^2-\bar{Q}_1-\bar{Q}_2-\bar{A})=\bar{\Omega}^*(\bar{P}^2-\bar{Q}_1-\bar{Q}_2-\bar{A}).$$

Hence we conclude that

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)=\bar{\pi}^*(\bar{P}^2-\bar{Q}_1-\bar{Q}_2-\bar{A}+v).$$

Now $v(M(\bar{P},\bar{A},\bar{P}^2))=v(\bar{y}/\bar{x})=v(x)-v(x^{-1}+y)=2$. Hence by [A4, Corollary of Theorem 2], we have

$$\bar{\pi}^*(\bar{P}^2-\bar{Q}_1-\bar{Q}_2-\bar{A}+v)\approx\bar{\pi}^*(P^1-P_1-P_2-(P_3,2)).$$

Finally, by [A7, Proposition 6 of Section 3] we have

$$\bar{\pi}^*(P^1 - P_1 - P_2 - (P_3, 2)) = \bar{\pi}^*(P^1 - P_1 - P_2) = \bar{Z}^p$$

in case $p = 2$; and by [A7, Proposition 3 of Section 3] we know that $\bar{\pi}^*(P^1 - P_1 = P_2 - (P_3, 2))$ is infinite and unsolvable in case $p \neq 2, 3, 5$.

4. A conic and two lines.

THEOREM 4. *Assume that W consists of an irreducible conic H and two distinct lines A_1 and A_2. Then $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W)$. If A_1 and A_2 are not simultaneously tangent to H, then $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W) \approx \bar{Z}^p \oplus \bar{Z}^p$. If A_1 and A_2 are both tangent to H, then in case of $p = 2$ we have $\bar{\pi}'(P^2 - W) = \bar{Z}^p \oplus \bar{Z}^p$; and in case of $p \neq 2$ we have an exact sequence*

$$0 \to \bar{\pi}'(P^2 - W) \to \bar{\pi}^*(P^1 - P_1 - P_2 - (P_3, 2)) \oplus \bar{Z}^p \to Z_2 \to 0$$

where P_1, P_2, P_3 are distinct points on a projective line P^1 over k; furthermore $\bar{\pi}'(P^2 - W)$ is infinite and unsolvable provided $p \neq 2, 3, 5$.

Proof. If A_1 and A_2 are not simultaneously tangent to H, then $\nu(H, W; P^2) \leqq 2$ and hence $\dim |H| - \nu(H, W; P^2) > 1$; consequently the assertion follows from [A3, Theorem 2]. Now assume that A_1 and A_2 are both tangent to H. Then $\nu(H, W; P^2) = 4$ and we cannot apply [A3, Theorem 2]. Let $Q_j = H \cap A_j$, and chose projective coordinates X, Y, Z in P^2 such that $Q_1: X = Y = 0$, $Q_2: X = Z = 0$; $A_1: Y = 0$, $A_2: Z = 0$. Then H is given by $X^2 + aYZ = 0$, $0 \neq a \in k$. Let $\bar{P}^2$ be a projective plane with coordinates $\bar{X}, \bar{Y}, \bar{Z}$ and let τ be the birational map of P^2 onto $\bar{P}^2$ given by[1]:

$$\bar{X} : \bar{Y} : \bar{Z} = XY : X^2 : YZ; \qquad X : Y : Z = \bar{X}\bar{Y} : \bar{X}^2 : \bar{Y}\bar{Z}.$$

Let P^1 be the line $\bar{X} = 0$; let P_1, P_2, P_3 be the distinct points on P^1 given respectively by $\bar{X} = \bar{Y} + a\bar{Z} = 0$, $\bar{X} = \bar{Z} = 0$, $\bar{X} = \bar{Y} = 0$. Let $\bar{A}_1$ be the line $\bar{Y} + a\bar{Z} = 0$, let $\bar{A}_2$ be the line $\bar{Z} = 0$, and let $\bar{A}_3$ be the line $\bar{Y} = 0$. Then Q_1, Q_2 are the only fundamental points of τ, $\tau(Q_1) = P^1$, $\tau(Q_2) = \bar{A}_3$; P_2 and P_3 are the only fundamental points of τ^{-1}, $\tau^{-1}(P_2) = A_1$, $\tau^{-1}(P_3) = (X = 0)$. Also $\tau[H] = \bar{A}_1$ and $\tau[A_2] = \bar{A}_2$. Let v be the valuation of $k(P^2)/k$ having center $(X = 0)$ on P^2. Then v has center P_3 on $\bar{P}^2$ and it can easily be checked that $v(M(P_3, P^1, \bar{P}^2) = 1$, and $v(M(P_3, \bar{A}_3, \bar{P}^2)) = 2$. Now $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ have the point $\bar{Y} = \bar{Z} = 0$ in common and it is not fundamental for τ^{-1}. Hence by [A4, Lemma 6 of Section 2], it follows that

$$\bar{\Omega}'(P^2 - W) \subset \bar{\Omega}'(\bar{P}^2 - \bar{A}_1 - \bar{A}_2 - \bar{A}_3 - P^1).$$

By [A4, Theorem 3 of Section 4],

$$\bar{\Omega}'(P^2-\bar{A}_1-\bar{A}_2-\bar{A}_3-P^1)=\bar{\Omega}^*(P^2-\bar{A}_1-\bar{A}_2-\bar{A}_3-P^1).$$

Therefore

$$\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)=\bar{\pi}^*(P^2-\bar{A}_1-\bar{A}_2-\bar{A}_3-P^1+v).$$

Now by [A4, Theorem 4 of Section 4], there exists an exact sequence

$$0\to\bar{\pi}^*(P^2-\bar{A}_1-\bar{A}_2-\bar{A}_3-P^1+v)$$
$$\to\pi^*(P^1-P_1-P_2-(P_3,2))\oplus\bar{Z}^p\to G\to 0$$

where $G=Z_1$ or Z_2 according as $p=2$ or $p\neq 2$. The rest follows from [A7, Propositions 3 and 6 of Section 3].

5. Four lines.

THEOREM 5. *Assume that W consists of four distinct lines. Then we have the following possibilities: (1) No three of the lines have a point in common; (2) three of the lines have a point in common and the fourth line does not pass through this point; (3) all the four lines have a point in common. In Case 1 we have $\pi'(P^2-W)=\bar{\pi}^*(P^2-W)=\bar{Z}^p\oplus\bar{Z}^p\oplus\bar{Z}^p$. In Case 2 we have $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\pi^*(P^1-P_1-P_2-P_3)\oplus\bar{Z}^p$, where P^1 is a projective line over k and P_1, P_2, P_3 are distinct points of P^1. In Case 3 we have $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{\pi}^*(P^1-P_1-P_2-P_3-P_4)$, where P^1 is a projective line over k and P_1, P_2, P_3, P_4 are distinct points on P^1. In Cases 2 and 3, $\bar{\pi}'(P^2-W)$ is infinite and unsolvable provided $p\neq 2,3,5$.*

Proof. Follows from [A1, Section 13], [A4, Section 4] and [A7, Section 3].

6. A conic and a line.

THEOREM 6. *Assume that W consists of an irreducible conic H and a line A. Then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)\approx\bar{Z}^p$.*

Proof. We have $\nu(H,W;P^2)\leqq 2$ and hence the result follows from [A3, Theorem 2].

7. Three lines.

THEOREM 7. *Assume that W consists of three distinct lines. If the three lines do not have a common point, then $\bar{\pi}'(P^2-W)=\bar{\pi}^*(P^2-W)$*

$\approx \bar{Z}^p \oplus \bar{Z}^p$. *If the three lines have a point in common, then* $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W) \approx \bar{\pi}^*(P^1 - P_1 - P_2 - P_3)$, *where* P^1 *is a projective line over* k *and* P_1, P_2, P_3 *are distinct point on* P^1; *consequently* $\bar{\pi}'(P^2 - W)$ *is infinite and unsolvable provided* $p \neq 2, 3, 5$.

Proof. Follows from [A1, Section 13], [A4, Section 4] and [A7, Section 3].

8. Two lines.

THEOREM 8. *Assume that* W *consists of two distinct lines. Then* $\bar{\pi}'(P^2 - W) = \bar{\pi}^*(P^2 - W) \approx \bar{Z}^p$.

Proof. Follows from [A1, Section 13].

PRINCETON UNIVERSITY.

REFERENCES.

S. Abhyankar,

[A1] "Tame coverings and fundamental groups of algebraic varieties. Part I: Branch loci with normal crossings," *American Journal of Mathematics*, vol. 81 (1959), pp. 46-94.

[A2] "———. Part II: Branch curves with higher singularities," *ibid.*, vol. 82 (1960), pp. 120-178.

[A3] "———. PartIII: Some other sets of condition for the fundamental group to be abelian," *ibid.*, vol. 82 (1960), pp. 179-190.

[A4] "———. Part IV: Product theorems," *ibid.*, vol. 82 (1960), pp. 341-364.

[A5] "———. Part V: Three cuspidal plane quartics," *ibid.*, vol. 82 (1960), pp. 365-373.

[A6] "Cubic surfaces with a double line," *Memoirs of the College of Science, University of Kyoto, Series A, Mathematics*, vol. 32 (1960), pp. 455-511.

[A7] "Coverings of algebraic curves," *American Journal of Mathematics*, vol. 79 (1957), pp. 825-856.

SEMINUCLEAR EXTENSIONS OF GALOIS FIELDS.*

By D. R. HUGHES [1] and ERWIN KLEINFELD.[2]

We consider here the problem of obtaining all division rings R (not associative) that are quadratic extensions of a Galois field F, where F is to be contained in the right and middle nuclei of R. This problem was inspired by the discovery of such a division ring with 16 elements ([2]). As a result we obtain a new class of finite division rings and a corresponding class of projective planes (in another paper ([1]) the collineation groups for this class of planes are discussed). We find that every Galois field not a prime field is capable of being extended in this day; if we further restrict R in such a way that F is to be contained in the nucleus of R, then exactly those Galois fields which are themselves quadratic extensions permit such an extension.

Definition. The *left nucleus* of a ring R is the set of all a in R such that $(ax)y = a(xy)$ for all x, y in R. The *middle nucleus* of R is the set of all b in R such that $(xb)y = x(by)$ for all x, y in R. The *right nucleus* of R is the set of all c in R such that $(xy)c = x(yc)$ for all x, y in R. The *nucleus* of R is the intersection of the right, middle and left nuclei.

From now on we shall assume that R is a not associative division ring that is a quadratic extension of a Galois field F, such that F is contained in the right and middle nuclei of R. Then R can be represented as a two-dimensional right vector space over F; let 1, λ be a basis of R over F. All elements of R have the form $x + \lambda y$, where x, y are in F; R will be completely determined once its multiplication is specified. If z is an arbitrary generator of the multiplicative group of non-zero elements of F, then the multiplication in R will be determined once $z\lambda$ and λ^2 are known. For we expand

$$(x + \lambda y)(u + \lambda v) = xu + \lambda(yu) + (x\lambda)v + \lambda(y\lambda)v.$$

If $\lambda^2 = \delta_0 + \lambda\delta_1$ and for any s in F, $s\lambda = s_0 + \lambda s_1$, then

* Received May 10, 1959.

[1] Supported in part by the United States Air Force under Contract No. AF 18(600)-1383.

[2] Supported in part by the United States Army Office of Ordnance Research.

$$\begin{aligned}(x+\lambda y)(u+\lambda v) &= xu+\lambda(yu)+x_0v+\lambda(x_1v)+\lambda[y_0v+\lambda(y_1v)]\\ &= (xu+x_0v+\delta_0y_1v)+\lambda(yu+x_1v+y_0v+\delta_1y_1v).\end{aligned}$$

Let us suppose that $z\lambda=q+\lambda r$ for the element z selected above. Two cases arise naturally: either $r=z$ or $r\neq z$. If $r=z$, then $z\lambda=q+\lambda z$; assume inductively that $z^i\lambda=iz^{i-1}q+\lambda z^i$. Then

$$z^{i+1}\lambda=z(z^i\lambda)=iz^iq+(z\lambda)z^i=iz^iq+z^iq+\lambda z^{i+1}=(i+1)z^iq+\lambda z^{i+1},$$

and so the formula is established. Suppose that F has p^n elements, where p is a prime, and substitute $i=p^n$ in the formula. Since then $z^{p^n}=z$, we have $z\lambda=\lambda z$, and so $q=0$. But then one verifies easily that R is commutative, and that λ is in the nucleus of R. In other words, R must be associative, contrary to assumption; so the case $r=z$ does not arise. We may therefore assume that $z\lambda=q+\lambda r$, where $z\neq r$. Now let $\lambda'=\lambda+q/(r-z)$. It is immediate that $z\lambda'=\lambda' r$ and therefore $z^j\lambda'=\lambda' r^j$. In other words, we could have selected λ in such a way that for every x in F, $x\lambda=\lambda x^\sigma$, where x^σ is in F. Let us examine the mapping σ in more detail. Since $(x+y)\lambda=x\lambda+y\lambda=\lambda x^\sigma+\lambda y^\sigma=\lambda(x^\sigma+y^\sigma)$, we see that $(x+y)^\sigma=x^\sigma+y^\sigma$. Also,

$$(xy)\lambda=x(y\lambda)=x(\lambda y^\sigma)=(x\lambda)y^\sigma=(\lambda x^\sigma)y^\sigma=\lambda(x^\sigma y^\sigma),$$

and so $(xy)^\sigma=x^\sigma y^\sigma$. Since R is a division ring, σ must be one-to-one, and so σ is an automorphism of F.

Multiplication in R is now completely determined by σ, and the elements δ_0 and δ_1 in F. In fact, we have

$$(1)\qquad (x+\lambda y)(u+\lambda v)=(xu+\delta_0y^\sigma v)+\lambda(yu+x^\sigma v+\delta_1y^\sigma v).$$

Two issues remain to be clarified. Namely, under what circumstances is R both not associative and a division ring, and under what circumstances is F contained in the right and middle nuclei of R? We answer the latter question first.

Let $a=x+\lambda y$, $b=u+\lambda v$, $c=w+\lambda z$ be three arbitrary elements of R. Then the associator $(a,b,c)=(ab)c-a(bc)$ may be easily calculated, using (1), and we find that

$$(2)\qquad \begin{aligned}(a,b,c) &= \delta_0v^\sigma z[(x^{\sigma^2}-x)+(\delta_1{}^\sigma y^{\sigma^2}-\delta_1y^\sigma)]\\ &\quad+\lambda v^\sigma z[(\delta_0{}^\sigma y^{\sigma^2}-\delta_0y)+(\delta_1{}^\sigma y^{\sigma^2}-\delta_1y^\sigma)+\delta_1(x^{\sigma^2}-x^\sigma)].\end{aligned}$$

If either $v=0$ or $z=0$ in (2), then $(a,b,c)=0$, and so F is certainly contained in the right and middle nucleus of R.

Now we investigate the conditions under which R will be a division ring. Since R is finite, R will be a division ring if and only if R has no divisors

of zero. Suppose $(x+\lambda y)(u+\lambda v)=0$, where not both u and v are zero. From (1) we obtain $xu+\delta_0 y^\sigma v=0$, as well as $yu+x^\sigma v+\delta_1 y^\sigma v=0$. Then the matrix of the two equations involving the variables u and v must be singular. The determinant is easily computed to be $x^{1+\sigma}+\delta_1 xy^\sigma-\delta_0 y^{1+\sigma}$, so we set this equal to zero. If $y=0$, then $x^{1+\sigma}=xx^\sigma=0$, so $x=0$. Therefore assume that $y\neq 0$, and let $w=xy^{-1}$; then $x^{1+\sigma}+\delta_1 xy^\sigma-\delta_0 y^{1+\sigma}$ $=y^{1+\sigma}[w^{1+\sigma}+\delta_1 w-\delta_0]=0$, and so $w^{1+\sigma}+\delta_1 w-\delta_0=0$. We have demonstrated that R is a division ring if and only if

$$w^{1+\sigma}+\delta_1 w-\delta_0=0 \tag{3}$$

has no solution for w in F.

Let us assume that in addition to the previous conditions, F is even contained in the nucleus of R. Putting $y=0$ in (2), we obtain (a,b,c) $=\delta_0 v^\sigma z(x^{\sigma^2}-x)+\lambda v^\sigma z(x^{\sigma^2}-x^\sigma)\delta_1=0$. Consequently $\delta_0 v^\sigma z(x^{\sigma^2}-x)=0$. If $\delta_0{}^* = 0$, then $\lambda^2=\lambda\delta_1$ implies that R has divisors of zero, contrary to assumption, and so $\delta_0\neq 0$. But then $x^{\sigma^2}=x$ for all x in F, so that $\sigma^2=I$, the identity mapping. Furthermore, $v^\sigma z(x^{\sigma^2}-x^\sigma)\delta_1=0$, so either $\sigma=I$ or $\delta_1=0$. But putting $\sigma=I$ in (1) we see that R would then be associative. So if R is a not associative division ring and F is in its nucleus, then $\sigma^2=I$ and $\delta_1=0$.

Suppose now that R is an associative division ring and $\sigma\neq I$. Then $\sigma^2=I$ and $\delta_1=0$. Substituting these values in (2), we discover that $\lambda v^\sigma yz(\delta_0{}^\sigma-\delta_0)=0$; consequently $\delta_0{}^\sigma=\delta_0$. Now if E is defined to be the fixed field of σ in F, then δ_0 is in E. As w ranges over F, $w^{1+\sigma}$ ranges over all of E, since σ has order two. Thus no matter how δ_0 is chosen, since $\delta_1=0$, (3) will have a solution for w in F, and hence R will not be a division ring. We have reached a contradiction. So if R is a division ring with F in the right and middle nuclei of R and R a quadratic extension of F, then R is associative if and only if $\sigma=I$.

We summarize the results obtained so far in the following theorems.

THEOREM 1. *Let R be a not associative division ring which is a quadratic extension of a Galois field F, and suppose F is contained in the right and middle nuclei of R. Then R must be isomorphic to a ring S constructed as follows: Let S be a vector space of dimension 2 over F, having basis 1, λ and multiplication defined by $(x+\lambda y)(u+\lambda v)=(xu+\delta_0 y^\sigma v)+\lambda(yu+x^\sigma v+\delta_1 y^\sigma v)$, where σ is an arbitrary non-identity automorphism of F and δ_0, δ_1 in F are subject only to the condition that $w^{1+\sigma}+\delta_1 w-\delta_0=0$ have no solution for w in F. Conversely, given F, σ, δ_0, δ_1, satisfying the above conditions, then S will satisfy the conditions on R.*

THEOREM 2. *Let R be a not associative division ring which is a quadratic extension of a Galois field F, and suppose F is contained in the nucleus of R. Then R must be isomorphic to one of the rings S of Theorem 1 with the additional stipulation that $\sigma^2 = I$ and $\delta_1 = 0$. Conversely, all such S satisfy the conditions on R.*

At this point the following question arises: given a Galois field F, does there exist an extension R satisfying Theorems 1 and 2? First we consider Theorem 1. In order to obtain such an R we need to produce an automorphism $\sigma \neq I$ of F and elements δ_0, δ_1 in F such that (3) has no solution in F. Suppose F has p^n elements, p a prime. If $n = 1$, $\sigma \neq I$ cannot exist, so R does not exist either. Assume that $n > 1$; two cases arise. If $p > 2$, choose $\delta_1 = 0$, and $x^\sigma = x^p$. Since $(-1)^{1+\sigma} = (-1)^{1+p} = 1 = (1)^{1+\sigma}$, there must exist an element not of the form $w^{1+\sigma}$, for w in F; let δ_0 be such an element. Then (3) is not satisfied by any w in F. If $p = 2$, choose $\delta_1 = 1$, $x^\sigma = x^2$. Since the mapping which sends x onto $x^3 + x$ send both 0 and 1 onto 0, there exists an element in F which is not of the form $x^3 + x$; choose δ_0 to be such an element. Again (3) cannot be satisfied by any w in F. Thus as long as F is not a prime field, an extension of F as described in Theorem 1 always exists.

A similar argument applies if an extension of F satisfying the hypotheses of Theorem 2 is to exist. In that case F must have an automorphism of order 2. So if F has order p^n, n must be even; that is, F must itself be a quadratic extension of a Galois field to begin with. Conversely, if F has p^{2k} elements then the extension described in Theorem 2 will always be possible by a suitable choice of σ and δ_0.

We conclude with the remark that while the construction of R using (1) will yield division rings when F is an infinite field, not all division rings which are quadratic extensions of such an F, with F in the middle and right nuclei, need be of the that form. In particular, Theorems 1 and 2 are no longer valid, and we have omitted the discussion of the infinite case.

UNIVERSITY OF CHICAGO,
OHIO STATE UNIVERSITY.

REFERENCES.

[1] D. R. Hughes, "Collineation groups of non-Desarguesian planes II. Some seminuclear division algebras," *American Journal of Mathematics*, vol. 82 (1960), pp. 113-119.

[2] Erwin Kleinfeld, "Construction of Veblen-Wedderburn planes," *Journal of the Association for Computing Machinery* (1960).

REGULAR MAPPINGS WHOSE INVERSES ARE 3-CELLS.*[1]

By Mary-Elizabeth Hamstrom.

1. Introduction. In our paper [6] Eldon Dyer and I introduced the notion of a completely regular mapping, f, of the metric space X onto a metric space Y (see Definitions 2.1 and 2.2) and were able to prove that under certain additional hypotheses on X, Y, and the inverses under f, (X, f, Y) is a locally trivial fibre space. Under some conditions, f is the projection mapping of a direct product. In [6] and [7] it was shown that if f is a 0-regular mapping of a metric space X onto a metric space Y and M is a compact 2-manifold with boundary such that each inverse under f is homeomorphic to M, then f is completely regular. Thus it may be proved that if X is complete and Y has finite covering dimension, then (X, f, Y) is a locally trivial fibre space and if Y is locally compact, separable, and contractible, then X is homeomorphic to the direct product $Y \times M$, f corresponding to the projection map of $Y \times M$ onto Y. The purpose of the present paper is to extend some of these results to the case where f is homotopy 2-regular and M is a compact 3-manifold with boundary which is imbeddable in E^3. A later paper will consider more general 3-manifolds [8].

Section 2 states some definitions and proves some lemmas concerning the convergence of the boundaries of the 3-manifold inverses under regular mappings. The principal result of Section 3 is Theorem 3.13, which states that if f is a homotopy 2-regular mapping of a metric space X onto a metric space Y such that each inverse under f is a 3-cell, M, then f is completely regular. The proof involves complicated constructions and push-pull arguments, which the reader may find easier to follow if he first reads Definition 2.7 and then the statements of the lemmas and theorems of Section 3 in decreasing order. Section 4 involves an induction argument on the number of elements in a cellular decomposition of M to yield Theorem 4.5, which extends Theorem 3.13 to the case where M is a compact 3-manifold with

* Received February 10, 1959.

[1] Presented to the American Mathematical Society, January 20, 1959. This work was started at the Institute for Advanced Study, when the author held National Science Foundation grants NSF-G2577 and NSF-G3964.

boundary and is imbeddable in E^3. Section 5 deals with the space of homeomorphisms on a 3-manifold and extends the results mentioned in the first paragraph above to completely regular mappings whose inverses are 3-manifolds with boundary which are imbeddable in E^3. Section 6 considers some slight relaxations in the hypotheses of some previous theorems.

2. Homotopy regular mappings whose inverses are 3-manifolds.

Definition 2.1. A proper mapping f of a metric space X onto a metric space Y is *homotopy n-regular* (*h-n-regular*) provided that it is true that f is open and if x is a point of X and ϵ is a positive number, then there is a positive number δ such that every mapping of a k-sphere, $k \leqq n$, into $S(x,\delta) \cap f^{-1}(y)$, $y \in Y$, is homotopic to 0 in $S(x,\epsilon) \cap f^{-1}(y)$. (A proper map is one for which inverses of compact sets are compact. The symbol $S(x,\epsilon)$ denotes the set of points with distance from x less than ϵ.)

Definition 2.2. The mapping f is said to be *completely regular* provided that if $\epsilon > 0$ and $y \in Y$, then there is a $\delta > 0$ such that $d(y, y') < \delta$, $y' \in Y$, implies that there is a homeomorphism of $f^{-1}(y)$ onto $f^{-1}(y')$ which moves no point as much as ϵ (i.e. an ϵ-homeomorphism).

In the sequel, a sequence $x_1, x_2, \cdots$ may sometimes be denoted by the symbol $\{x_i\}_i$ or, if no ambiguity results, by $\{x_i\}$. A point P with the property that every open set containing P intersects all but a finite number of the elements of the sequence $\{x_i\}$ will be called a *sequential limit point* of $\{x_i\}$.

Definition 2.3. If f is an h-n-regular (completely regular) mapping of a compact space X onto a space Y and Y consists of the points of the sequence $y_0, y_1, y_2, \cdots$ which converges to y_0, then the sequence $\{f^{-1}(y_i)\}$ is said to *converge h-n-regularly* (*completely regularly*) to $f^{-1}(y_0)$.

In the remainder of this section, $M_0, M_1, M_2, \cdots$ will denote the elements of a sequence of compact 3-manifolds with boundary converging h-2-regularly to M_0, the union of whose elements is a compact metric space. It will further be supposed that the boundaries, $K_0, K_1, K_2, \cdots$, of these 3-manifolds are mutually homeomorphic and that each M_i is polyhedrally imbeddable in E^3. The space $\cup M_i$ is clearly finite dimensional and hence may be considered as a subset of some Euclidean space, E^{n-3}. Furthermore, it follows from a result of Klee ([10], 3.3, p. 36) that $\cup M_i$ may be so imbedded in E^n that M_0 is a polyhedral subset of a 3-dimensional hyperplane of E^n.

The 3-manifold M_i may be triangulated by means of a triangulation Γ_i.

In each M_i, polyhedra, polygons, et cetera will be defined relative to Γ_i. No particular relationships among the Γ_i's will be assumed. Distances will be ordinary distances in E^n. *Unless it is explicitly stated otherwise, any subset of M_i that is referred to will be polyhedral.*

If M is a manifold with boundary, the notation bdry M and int M will be used to denote the sets of boundary points and non-boundary points of M respectively.

The content of the first two lemmas is probably known, but they are included for completeness.

LEMMA 2.4. *If M is a compact 2-manifold with boundary, which may be empty, and ϵ is a positive number, there is a positive number δ such that each mapping of M into itself which moves no point as much as δ is ϵ-homotopic to the identity map.*

Indication of proof. Dente by $J_1, J_2, \cdots, J_m$ the boundary simple closed curves of M and by $A_1, \cdots, A_m$ mutually exclusive annuli in M such that for each i, A_i is bounded by J_i and a simple closed curve J_i' which is $\epsilon/2$-isotopic to J_i in A_i. For sufficiently "narrow" A_i and small δ, each δ-mapping f of M into itself carries J_i into $A_i - J_i'$ and is thus ϵ-homotopic relative to $f^{-1}(M - A_i)$ to a mapping which, when restricted to J_i, is the identity. The proof of Lemma 2.4 may thus be restricted to the consideration of mappings leaving bdry M pointwise fixed.

Let G be a cellular decomposition of M in the sense that G is a finite collection $g_1, g_2, \cdots, g_k$ of discs such that $\cup g_i = M$ and if $g_i \cap g_j$ exists, it is an arc. The proof proceeds by induction on k. The lemma is clearly true if $k = 1$. Suppose it to be true for all compact 2-manifolds with boundary which have a cellular decomposition with fewer than k elements and that $G: g_1, \cdots, g_k$ is a cellular decomposition of M. It may be assumed that each component of $g_i \cap$ bdry M is an arc. If c is an element of G, $\mathrm{cl}(M - c)$ is a compact 2-manifold with boundary which has a cellular decomposition with fewer than k elements. A slight modification of the content of the previous paragraph may be used to prove that for sufficiently small δ, each δ-mapping of M onto itself leaving bdry M pointwise fixed is $\epsilon/2$-homotopic to a mapping leaving bdry $M \cup$ bdry c pointwise fixed. The induction hypothesis may now be used to prove the lemma.

LEMMA 2.5. *If f is a mapping of a compact 2-manifold M onto a compact 2-manifold N which is not homotopic to a mapping carrying M into a proper subset of N and A is an annulus in N, then some simple closed curve,*

J, in $f^{-1}(A)$ is mapped essentially into A (i.e. $f|J$ is not homotopic to 0 in A). As a consequence, for sufficiently small ϵ, if f^* is a piecewise linear ϵ-approximation to f and z is a non-trivial 1-cycle carried by J, then $f^*(z)$ does not bound in A.

Proof. Let A' and A'' denote annuli such that $A'' \subset \operatorname{int} A'$, $A' \subset \operatorname{int} A$ and each circles those in which it is contained, let K denote the closure of $A - A'$ and let t denote an arc in A' which lies except for its endpoints in $\operatorname{int} A'$ and whose endpoints lie in different components of K. Suppcse that no simple closed curve in $f^{-1}(A)$ has the required property. If U is a component of $f^{-1}(A) - f^{-1}(K)$ the closure of whose image lies in $A' - A''$, then f is homotopic relative to $M - U$ to a mapping carrying U into K and $f(U)$ remains in $A' - A''$ during the homotopy. Thus it may be assumed that $f^{-1}(A) - f^{-1}(K)$ has only a finite number of components. If U is one such component, there is a 2-manifold with boundary, N', lying in U such that the closure of $f(U - N')$ lies in $A' - A''$. Since $f | \operatorname{bdry} N'$ is homotopic to 0 in A, there is a 2-manifold with boundary, N'', in N' which is homeomorpic to N' and whose boundary is in a small neighborhood of $\operatorname{bdry} N'$ in $f^{-1}(A' - A'')$ and there is a homotopy of f relative to $M - N'$ into a mapping g carrying $\operatorname{bdry} N''$ into two points, $f(N' - N'')$ remaining in $A' - A''$ during the homotopy. The mapping g is, in turn, homotopic relative to $M - U$ to a mapping f_1 which carries U into $K \cup t$, $f(U - N')$ remaining $A' - A''$ during the homotopy. Repeat this process until a mapping f_k is obtained which is homotopic to f and carries M into $(M - A') \cup t$. This contradiction proves the lemma.

LEMMA 2.6. *If ϵ is a positive number and H_0 is a polyhedron in M_0, then there is an integer N such that if $i > N$, then there is a piecewise linear ϵ-mapping of H_0 into $\operatorname{int} M_i$.*

Proof. The compactness of $\cup M_i$ and the h-2-regularity of the convergence of $\{M_i\}$ to M_0 imply the existence of numbers $\delta_0, \delta_1, \delta_2, \delta_3 = \epsilon$, $0 < \delta_1 < \delta_2 < \delta_3$, such that every mapping of a k-sphere, $k \leqq 2$, into a subset of M_i of diameter less than δ_k is homotopic to 0 on a subset of M_i of diameter less than $\delta_{k+1}/(k+3)$. Let T_0 be a triangulation of H_0 of mesh less than $\delta_0/3$, the simplices of T_0 being in some subdivision of Γ_0. (The mesh of T_0 is the maximum of the diameters of the simplices of T_0.) There is an integer N such that if $i > N$, then there is a $\delta_0/3$-homeomorphism $g_i{}^0$ of the 0-skeleton, $T_0{}^0$ of T_0 into $\operatorname{int} M_i$. If a and b are the vertices of a 1-simplex, $s_0{}^1$, of T_0, $d(g_i{}^0(a), g_i{}^0(b)) < \delta_0$. Consequently, there is an arc

s_i^1 in M_i of diameter less than $\delta_1/3$ whose endpoints are $g_i^0(a)$ and $g_i^0(b)$. This arc may be taken to be polygonal and a subset of $\operatorname{int} M_i$. Also, small changes in the s_i^1's can be made so that no two of them intersect except, possibly, at their endpoints.

If each mapping $g_i^0 \mid (a \cup b)$ is extended to a homeomorphism of s_0^1 onto s_i^1, g_i^0 is extended to a piecewise linear homeomorphism g_i^1 of the 1-skeleton T_0^1 of T_0 into $\operatorname{int} M_i$. Clearly g_i^1 is a δ_1-homeomorphism and if ∂s_0^2 is the boundary of a 2-simplex s_0^2 of T_0, $\operatorname{diam} g_i^1(\partial s_0^2) < \delta_1$. Thus $g_i^1 \mid \partial s_0^2$ can be extended to a mapping of s_0^2 onto a subset s_i^2 of M_i of diameter less than $\delta_2/4$, which, by the simplicial approximation theorem, may be assumed to be piecewise linear. Small changes in s_i^2 may be made, if necessary, so that it lies in $\operatorname{int} M_i$. In this way, there is obtained a piecewise linear extension, g_i^2 of g_i^1 a δ_2-mapping of the 2-skeleton, T_0^2, of T_0 into $\operatorname{int} M_i$ with the property that if ∂s_0^3 is the boundary of a 3-simplex s_0^3 of T_0, then $\operatorname{diam} g_i^2(\partial s_0^3) < \delta_2$.

It follows as above that $g_i^2 \mid \partial s_0^3$ can be extended to a piecewise linear ϵ-mapping of s_0^3 into a subset s_i^3 of $\operatorname{int} M_i$. This defines a piecewise linear ϵ-mapping of H_0 into $\operatorname{int} M_i$.

A *regular neighborhood in* M_i of a polyhedron P_i is a closed neighborhood of which P_i is a (strong) deformation retract and whose boundary is a compact polyhedral 2-manifold (not necessarily connected and with boundary if P_i intersects K_i). It is known that every polyhedron in M_i has such a neighborhood (Whitehead, see [5] and [16]). A *regular neighborhood in* E^n of a polyhedron P in M_0 is a closed neighborhood U of which P and $U \cap M_0$ are deformation retracts, which is such that $U \cap M_0$ is a regular neighborhood of P in M_0. The mappings of U onto P and $U \cap M_0$ resulting from the deformation will be called the natural mappings of U onto P and $U \cap M_0$. For each positive number ϵ there exist such neighborhoods U for which no point is moved as much as ϵ during the deformations of U on $U \cap M_0$ and $U \cap M_0$ on P. Such neighborhoods will be called regular ϵ-neighborhoods.

If J is a subpolyderon of P, above, U' is a regular ϵ-neighborhood of J in E^n and r and q are the deformations of U into $U \cap M_0$ and $U \cap M_0$ into P, then U' is said to be *consistently imbedded* in U provided that (1) $U' \cap M_0$ and $(U - U') \cap M_0$ are deformation retracts of U' and $U - U'$ under r, (2) $U' \cap P$ and $(U - U') \cap P$ are deformation retracts of $U' \cap M_0$ and $(U - U') \cap M_0$ under q and (3) J is a deformation retract of $U' \cap P$.

Definition 2.7. If H_0 is a compact polyhedral p-manifold, $p = 1, 2$, in M_0 and $\{H_i\}$ is a sequence of compact polyhedral p-manifolds converging to

H_0 such that for each i, H_i lies in M_i, then the sequence $\{H_i\}$ is said to *converge strongly* to H_0 if it is true that for each regular neighborhood in E^n, V_0, of H_0 and for sufficiently large i, the p-cycles carried by H_i (mod the integers) fail to bound in $V_0 \cap M_i$. If $\{D_i\}$ is a sequence of discs (annuli) converging to the disc (annulus) D_0 in M_0 and for each i, $D_i \subset M_i$, then the sequence $\{D_i\}$ is said to converge strongly to D_0 if the sequence $\{\text{bdry}\, D_i\}$ converges strongly to bdry D_0.

All homology in this paper will be taken modulo the integers.

LEMMA 2.8. *The sequence $K_1, K_2, \cdots$ converges to a subset of K_0.*

Proof. Denote by C_0 a component of K_0 and by U_0 a regular neighborhood of C_0 in E^n, p the natural mapping. Presume that $U_0 \cap (K_0 - C_0) = 0$. It follows from Lemma 2.6 that there are a sequence $\epsilon_1, \epsilon_2, \cdots$ of positive numbers converging to 0 and a sequence $g_1, g_2, \cdots$ of mappings such that for each i, g_i is a piecewise linear ϵ_i-mapping of C_0 into $U_0 \cap \text{int}\, M_i$. For each i, a regular neighborhood V_i of $g_i(C_0)$ may be constructed in $U_0 \cap \text{int}\, M_i$ in such a way that the sequence $\{V_i\}$ converges to C_0. If V_i fails to separate M_i, then, since $V_i \cap K_i = 0$, the boundary of V_i has only one component and the 2-cycle carried by $g_i(C_0)$ into which a fundamental 2-cycle γ_0 of C_0 is mapped by g_i bounds in V_i. If V_i does separate M_i, it follows from the h-2-regularity of the convergence of $\{M_i\}$ to M_0 that for sufficiently large i all but one of the components of $M_i - V_i$ lies in U_0. Denote by V_i' the union of V_i and these components and by C_i' its boundary. If none of these components intersects K_i, then C_i' is a compact, connnected 2-manifold and the 2-cycle $g_i(\gamma_0)$ bounds in V_i'. In any case, $g_i(\gamma_0)$ bounds in $U_0 \cap \text{int}\, M_i$ and thus $p_i g_i(\gamma_0)$ bounds in C_0, where p_i is a piecewise linear ϵ_i-approximation to $p \mid V_i'$. It follows from Lemma 2.4 that for sufficiently large i, $p_i g_i$ is homotopic to the identity mapping of C_0 into itself and thus carries γ_0 into a non-bounding cycle. Since $g_i(\gamma_0)$ bounds in V_i', this is a contradiction. Hence, for sufficiently large i, V_i separates M_i and V_i' contains a component C_i of K_i. The sequence $\{C_i\}$ converges to a subset of C_0. Since this argument holds for each component of K_0 and K_i has the same number of components as K_0, the lemma is proved.

In the next several lemmas, C_0 denotes a component of K_0 and $C_1, C_2, \cdots$ denote components of $K_1, K_2, \cdots$ converging to a subset of C_0. If U_0 is a regular neighborhood of C_0 in E^n not intersecting $K_0 - C_0$, then for sufficiently large i, C_i is the only component of K_i in U_0 and is one of the two components of bdry V_i', the other being denoted by C_i' (V_i' is as defined above).

LEMMA 2.9. *The sequence $\{C_i\}$ converges strongly to C_0.*

Proof. Let p be the natural mapping of U_0 onto C_0 and p_i a piecewise linear ϵ_i-approximation to $p \mid V_i'$, g_i and ϵ_i being defined as in the proof of Lemma 2.8. If γ_i and γ_i' are fundamental 2-cycles carried by C_i and C_i', γ_i is homologous to γ_i' in V_i' (γ_i' being taken with appropriate orientation) and thus $p_i(\gamma_i)$ is homologous to $p_i(\gamma_i')$ in C_0. Hence if γ_i bounds in $U_0 \cap M_i$, then $p_i(\gamma_i)$ and $p_i(\gamma_i')$ are homologous to 0 in C_0. However, $g_i(\gamma_0)$ is a linear combination of γ_i and γ_i'. Thus $p_i g_i(\gamma_0)$ bounds in C_0, contradicting the fact that for sufficiently large i, $p_i g_i$ is homotopic to the identity map of C_0 onto itself and thus carries the non-bounding cycle γ_0 into a non-bounding cycle. Therefore, γ_i fails to bound in $U_0 \cap M_i$.

If a subsequence $\{C_{n_i}\}_i$ of $\{C_i\}$ converges to a proper subset of C_0, then for sufficiently large i, $p_{n_i}(\gamma_{n_i})$ bounds in C_0 and a contradiction follows as above. This completes the proof that the convergence of $\{C_i\}$ to C_0 is strong.

LEMMA 2.10. *If J_0 is a simple closed curve in C_0, there is a sequence $J_1, J_2, \cdots$ of simple closed curves converging strongly to J_0 such that for each i, J_i is a subset of C_i.*

Proof. Let W_0 denote a regular neighborhood of J_0 in E^n consistently imbedded in U_0. If $p \mid C_i$ is homotopic to a map of C_i onto a proper subset of C_0, then $p_i(\gamma_i)$ bounds in C_0, which the proof of Lemma 2.9 has demonstrated to be false. Thus Lemma 2.5 may be applied to yield Lemma 2.10.

LEMMA 2.11. *If, for sufficiently large i, the fundamental 1-cycle, z_i, carried by J_i bounds in C_i, then z_0, the fundamental 1-cycle carried by J_0, bounds in C_0.*

Proof. For sufficiently large i, a piecewise linear approximation to $p \mid C_i$ carries z_i into a bounding cycle in C_0 which is a multiple of z_0. Thus z_0 bounds in C_0.

LEMMA 2.12. *If z_0 bounds in C_0, then for sufficiently large i, z_i bounds in C_i.*

Proof. Denote by $z_{01}, \cdots, z_{0h}$ a linearly independent system of cycles generating the first homology group, $H^1(C_0)$ of C_0, z_{0j} being carried by a simple closed curve J_{0j} which does not meet J_0, and for each j, let $\{J_{ij}\}_i$ be a sequence of simple closed curves converging strongly to J_{0j}, J_{ij} lying in C_i and carrying a fundamental 1-cycle z_{ij}. A modification of the proof of Lemma 2.11 demonstrates that for sufficiently large i, the z_{ij} are linearly independent and thus that the first Betti number, $p^1(C_i)$, of C_i is not less than $p^1(C_0)$.

However, K_i is homeomorphic to K_0 and this argument can be applied to each component of K_0. Thus $p^1(C_i) = p^1(C_0)$ for sufficiently large i and C_i is homeomorphic to C_0. If z_i does not bound in C_i then a multiple of z_i carried by J_i is a linear combination of $z_{i1}, \cdots, z_{ih}$, $cz_i = \sum c_j z_{ij}$, where not all the c_j are 0. Thus, applying the proof of Lemma 2.11, a multiple of z_0 is a (non-trivial) linear combination of $z_{01}, \cdots, z_{0h}$. This is impossible. Thus, for sufficiently large i, z_i bounds in C_i.

LEMMA 2.13. *If B_0 is a disc with boundary J_0 such that $B_0 - J_0 \subset \operatorname{int} M_0$ and $J_0 \subset K_0$, then there is a sequence $B_1, B_2, \cdots$ of discs, with boundaries $J_1, J_2, \cdots$, converging strongly to B_0 such that for each i, $B_i - J_i \subset \operatorname{int} M_i$ and $J_i \subset K_i$.*

Proof. Denote by ϵ a positive number and by U_0 and W_0 regular ϵ-neighborhoods of B_0 and J_0 in E^n, W_0 being consistently imbedded in U_0 and p and q denoting the natural mappings of U_0 and W_0 onto B_0 and J_0. There are positive numbers δ and $\delta' < \delta$ such that every (singular) 1-sphere in M_i of diameter less than δ bounds a (singular) 2-cell in M_i of diameter less than $\epsilon/2$ and every 0-phere in M_i of diameter less than δ' bounds a (singular) 1-cell in M_i of diameter less than $\delta/4$.

It follows from Lemma 2.10 that there is a sequence $\{J_i\}$ of simple closed curves converging strongly to J_0 such that for each i, J_i lies in K_i. There is a simple closed curve J_0' bounding a disc B_0' which is $\delta'/16$-homeomorphic to B_0 and lies in $\operatorname{int} B_0$. For sufficiently large i, $q \mid J_i$ is a $\delta'/16$-mapping q_i of J_i onto J_0 and there is a piecewise linear $\delta'/16$-mapping, g_i, of B_0' into $\operatorname{int} M_i \cap U_0$. Therefore there exists a piecewise linear $\delta'/4$-mapping h_i of J_i onto $g_i(J_0')$. Let $P_{i0}, \cdots, P_{im}$ be a sequence of points of J_i in that order such that the diameter of each component of $J_i - \cup P_{ij}$ is less than $\delta'/4$. There is an arc t_{ij} in M_i of diameter less than $\delta/4$ whose endpoints are P_{ij} and $h_i(P_{ij})$ and which may be constructed so as to meet K_i only in P_{ij}. The (singular) closed curve $t_{ij} \cup t_{ij-1} \cup P_{ij-1}P_{ij} \cup h_i(P_{ij-1}P_{ij})$ has diameter less than δ and consequently bounds a (singular) 2-cell B_{ij} of diameter less than $\epsilon/2$ which lies in $U_0 \cap M_i$ and may be constructed so as to meet K_i only in the arc $P_{ij-1}P_{ij}$ of J_i. The curve $h_i(J_i)$ is contractible in $g_i(B_0')$ and therefore the (singular) 2-cells B_{ij}, $j = 1, \cdots, m$, and $g_i(B_0')$ may be fitted together to form a singular 2-cell in $U_0 \cap M_i$ which is bounded by J_i, meets K_i only in J_i and has no singularities on its boundary. Hence it follows from Dehn's Lemma [12] that J_i bounds a non-singular 2-cell in $U_0 \cap M_i$ which meets K_i only in J_i. Since the sequence $\{J_i\}$ converges strongly to J_0 and U_0 is arbitrary, the existence of the required sequence is demonstrated.

THEOREM 2.14. *The sequence $\{K_i\}$ converges to K_0 h-1-regularly.*

Proof. Consider a component C_0 of K_0 and a sequence $\{C_i\}$ of components of $K_1, K_2, \cdots$ converging strongly to C_0 and all homeomorphic to C_0. Denote by P_0 a point of C_0, by J_0 a simple closed curve bounding a disc A_0 in C_0 whose interior contains P_0 and by B_0 the closure of $C_0 - A_0$. Let E_0 be a disc in M_0 such that $K_0 \cap E_0 = \operatorname{bdry} E_0 = J_0$. It follows from Lemma 2.13 that there are a sequence of simple closed curves $J_1, J_2, \cdots$ converging strongly to J_0 and a sequence $E_1, E_2, \cdots$ of discs converging strongly to E_0 such that for each i, $E_i \cap K_i = E_i \cap C_i = \operatorname{bdry} E_i = J_i$ and $E_i \subset M_i$. It follows readily from Lemma 2.12 that for each i, J_i separates C_i into two sets A_i and B_i whose closures are 2-manifolds with boundary and E_i separates M_i into two sets U_i and V_i, U_i bounded by $E_i \cup A_i$, V_i bounded by $E_i \cup B_i$. It follows from the 2-regularity of the convergence of $\{M_i\}$ to M_0 that the A_i and B_i may be so named that $\{A_i\}$ converges to A_0 and $\{B_i\}$ converges to B_0. Thus, since B_0 contains all the handles of C_0, it follows from Lemma 2.11 and the proof of Lemma 2.12 that B_i contains all the handles of C_i, for sufficiently large i. Therefore A_i is a disc.

If ϵ is a positive number, choose A_0 to lie in $S(P_0, \epsilon)$ and δ to be such that $S(P_0, \delta) \cap C_i \subset A_i$. Since for sufficiently large i, $A_i \subset S(P_0, \epsilon)$ and $B_i \subset C_i - (C_i \cap S(P_0, \delta))$, it follows that every i-sphere $(i = 0, 1)$ in $S(P_0, \delta) \cap C_i$ bounds an $(i+1)$-cell in $S(P^0, \epsilon) \cap C_i$ and the theorem is proved.

COROLLARY. *The sequence $K_1, K_2, \cdots$ converges to K_0 completely regularly.*

Proof. This is a direct consequence of Lemma 3 of [7] mentioned in the introduction.

LEMMA 2.15. *If S_0 is a 2-sphere bounding a 3-cell A_0 in $\operatorname{int} M_0$, then there is a sequence $S_1, S_2, \cdots$ of 2-spheres converging strongly to S_0 such that for each i, S_i bounds a 3-cell A_i in M_i and the sequences $\{A_i\}$ and $\{M_i - A_i\}$ converge to A_0 and $\operatorname{cl}(M_0 - A_0)$.*

Proof. Let U denote a regular ϵ-neighborhood of S_0 in $E^n - K_0$ and U' and V regular ϵ-neighborhoods of S_0 in $E^n - K_0$ such that $V \subset \operatorname{int} U$, $U \subset \operatorname{int} U'$, V is a deformation retract of U and U is a deformation retract of U'. Let p denote the natural mapping of U' onto S_0. There exist, as a consequence of Lemma 2.6, a sequence $\{\delta_i\}$ of positive numbers converging to 0 and sequences $\{g_i\}$ and $\{g_i'\}$ such that for each i, g_i is a piecewise-linear

δ_i-mapping of M_0 into $\operatorname{int} M_i$ and g_i' is a piecewise-linear δ_i-mapping of M_i into $\operatorname{int} M_0$. For sufficiently large i, $\operatorname{int} U \cap M_i$ is a 3-manifold in M_i which contains $g_i(V)$, $g_i'g_i(V)$ is in $\operatorname{int} U$, $g_i'(U \cap M_i)$ is in $\operatorname{int} U'$, and $g_i'(K_i)$ is in $(M_0 - A_0) - ((M_0 - A_0) \cap U')$.

Suppose that $g_i \mid S_0$ is homotopic to 0 in $\operatorname{int} U \cap M_i$ for infinitely many values of i, so that $g_i \mid S_0$ can be extended to a mapping G_i of A_0 into $U \cap M_i$. Then $g_i'G_i$ is a mapping of A_0 into $\operatorname{int} U'$, which implies that $g_i'g_i \mid S_0$ is homotopic to 0 in U', which, for sufficiently large i and small δ_i, is impossible. Hence, for sufficiently large i, $g_i \mid S_0$ is not homotopic to 0 in $U \cap M_i$. It follows from the work of C. D. Papakyriakopoulos [12] (the Sphere Problem) that arbitrary small neighborhoods of $g_i(S_0)$ contain polyhedral 2-spheres, S_i, which are not contractible in $U \cap M_i$. Since this construction can be made for each ϵ, the existence of a sequence $\{S_i\}$ of 2-spheres converging strongly to S_0 is demonstrated. If a non-trivial 2-cycle γ_i carried by S_i bounds in $U \cap M_i$, then, since M_i is imbeddable in E^3, S_i bounds a 3-cell in $U \cap M_i$ and is thus contractible therein, which is impossible.

The 2-regularity of the convergence of $\{M_i\}$ to M_0 implies that for sufficiently large i, one of the components of $M_i - S_i$ contains K_i. The closure of the other is, therefore, a 3-cell, A_i. The 2-regulatory also implies that every point of $M_0 - A_0$ is a sequential limit point of $\{M_i - A_i\}$ and no subsequence of $\{A_i\}$ converges to a subset of U or has a sequential limit point in $M_0 - A_0$. Thus $\{A_i\}$ converges to A_0 and $\{M_i - A_i\}$ converges to $\operatorname{cl}(M_0 - A_0)$, which was to be proved.

LEMMA 2.16. *The notation being that of Lemma* 2.15, *if T_0 is a torus in* $\operatorname{int} A_0$ *with interior U_0, then there is a sequence $T_1, T_2, \cdots$ of compact 2-manifolds converging strongly to T_0 such that for each i, T_i lies in* $\operatorname{int} A_i$ *with interior U_i and the sequences $\{U_i\}$ and $\{M_i - U_i\}$ converge to $T_0 \cup U_0$ and* $\operatorname{cl}(M_0 - U_0)$.

Proof. Let $\epsilon_1, \epsilon_2, \cdots$ be a sequence of positive numbers converging to 0 and $g_1, g_2, \cdots$ a sequence of mappings such that for each i, g_i is a piecewise linear ϵ_i-mapping of T_0 into $\operatorname{int} A_i$. There is a regular neighborhood V_0 of T_0 in E^n such that $V_0 \cap M_0 \subset \operatorname{int} A_0$ and for each i, there is a regular neighborhood V_i of $g_i(T_i)$ in $\operatorname{int} A_i$. The V_i may be so selected that $\{V_i\}$ converges to T_0. Denote the natural mapping of V_0 on T_0 by p.

The h-2-regularity of the convergence of $\{M_i\}$ to M_0 implies that for sufficiently large i, all but at most two of the components of $M_i - V_i$ lie in V_0. Denote the union of V_i and the components of $M_i - V_i$ in V_0 by N_i and the boundary of the component of $M_i - N_i$ containing K_i by T_i. If

$M_i - N_i$ were connected, a fundamental 2-cycle, γ_i, carried by T_i would bound in $V_0 \cap M_i$. Hence $g_i(\gamma_0)$, γ_0 being a fundamental 2-cycle carried by T_0, would bound in $V_0 \cap M_i$, since it lies in N_i. Hence $p_i g_i(\gamma_0)$ would bound in T_0, where p_i is a piecewise linear ϵ_i-approximation to $p \mid N_i$. But it follows from Lemma 2.4 that for sufficiently large i and small ϵ_i, $p_i g_i$ is homotopic to the identity map, so $p_i g_i(\gamma_0)$ does not bound. This contradiction implies that $M_i - N_i$ has two components, the one containing K_i and bounded *by* T_i, the other denoted by U_i' and bounded by T_i', which carries a fundamental 2-cycle γ_i'.

Suppose that γ_i bounds in $V_0 \cap M_i$. Then $p_i(\gamma_i)$ bounds in T_0 and consequently so does $p_i(\gamma_i')$. But $g_i(\gamma_0)$ is a linear combination of γ_i and γ_i'. Hence $p_i g_i(\gamma_0)$ bounds in T_0 for sufficiently large i—*a* contradiction. Also, if a subsequence $\{T_{n_i}\}$ of $\{T_i\}$ converges to a proper subset of T_0, $p_{n_i}(\gamma_{n_i})$ bounds in T_0 and another contradiction arises.. Thus $\{T_i\}$ converges strongly to T_0.

Denote by U_i the interior of $N_i \cup U_i'$. If $T_i \cup U_i$ lies in V_0, then γ_i bounds in $V_0 \cap M_i$ and $p_i(\gamma_i)$ bounds in T_0 for sufficiently large i. This contradiction and the 2-regularity of the convergence of $\{M_i\}$ to M_0 imply that $\{T_i \cup U_i\}$ converges to $T_0 \cup U_0$ and $\{M_i - U_i\}$ converges to $M_0 - U_0$.

3. **Regular mappings whose inverses are 3-cells.** In this section, M_i denotes a 3-cell and K_i its boundary. The remainder of the notation is that of Section 2. The 3-cell M_0 may be assumed to consist of those points $(x_1, \cdots, x_n)$ of E^n for which $x_1^2 + x_2^2 + x_3^2 = 1$ and $x_i = 0$ for $4 \leqq i \leqq n$. It follows from results in Section 2 that are sequences $\{g_i\}$ and $\{g_i'\}$ of mappings and a sequence $\{\epsilon_i\}$ of positive numbers converging to 0 such that for each i, g_i and g_i' are piecewise linear ϵ_i-mappings M_0 onto M_i and M_i onto M_0 such that $g_i \mid K_0$ and $g_i \mid K_i$ are homeomorphisms and $g_i(M_0 - K_0) = M_i - K_i = g_i'^{-1}(M_0 - K_0)$.

Definition 3.1. If M is a 3-cell bounded by a 2-sphere K and t is an arc lying except for its endpoints, which lie in K, in $\operatorname{int} M$, then t is said to be *unknotted* in M provided that there is a piecewise linear homomorphism carrying M onto a (solid) cube S in E^3 and carrying t onto a straight line interval in S. If A is an annulus except for its boundary, which lies in K, in $\operatorname{int} M$, then A is said to be unknotted in M provided that there is a piecewise linear homeomorphism of M onto S which carries A onto the union of all intervals in S intersecting a triangle on $\operatorname{bdry} S$ and parallel to a fixed interval in S.

LEMMA 3.2. *Suppose that* t_0 *is an unknotted polyhedral arc in* M_0 *with endpoints* P_0' *and* P_0'' *such that* $t_0 \cap K_0 = P_0' \cup P_0''$. *Then there is a sequence* $t_1, t_2, \cdots$ *of arcs converging to* t_0 *such that* (1) *for each* i, t_i *is unknotted in* M_i *and has endpoints* P_i' *and* P_i'' *and* (2) *the sequences* $\{P_i'\}$ *and* $\{P_i''\}$ *converge to* P_0' *and* P_0''.

Proof. There is a disc D_0 in M_0 with boundary J_0 such that $t_0 \subset D_0$ and $D_0 \cap K_0 = J_0$. It follows from Lemma 2.13 and the corollary to Theorem 2.14 that there is a sequence $D_1, D_2, \cdots$ of discs converging to D_0 such that for each i, $D_i \subset M_i$, $D_i \cap K_i = \operatorname{bdry} D_i = J_i$ and the sequences $\{J_i\}$ converges h-0-regularly to J_0. Let E_0 be a disc which is the closure of one of the components of $K_0 - J_0$ and let $E_1, E_2, \cdots$ be a sequence of discs converging to E_0 such that for each i, E_i is the closure of a component of $K_i - J_i$. Then the sequence of 2-spheres, $\{E_i \cup D_i\}$ converges strongly to $E_0 \cup D_0$. If s_0 is an arc in E_0 such that $s_0 \cap J_0 = P_0' \cup P_0''$, it follows from Lemma 2.5, applied to the natural mapping of a regular neighborhood of $E_0 \cup D_0$ onto $E_0 \cup D_0$ that there is a sequence $C_1, C_2, \cdots$ of simple closed curves converging strongly to $s_0 \cup t_0$ such that for each i, $C_i \subset D_i \cup E_i$. If ϵ is a positive number, then for sufficiently large i, there is an arc t_i in $C_i \cap D_i$ with endpoints P_i' and P_i'' such that $t_i \cap J_i = P_i' \cup P_i''$, $d(P_i', P_0') < \epsilon$ and $d(P_i'', P_0'') < \epsilon$. The sequence $\{t_i\}$ is the required sequence.

LEMMA 3.3. *Suppose that* A_0 *is an annulus in* M_0 *bounded by the simple closed curves* J_0' *and* J_0'' *such that* $A_0 \cap K_0 = J_0' \cup J_0''$ *and that* U_0 *and* V_0 *are the components of* $M_0 - A_0$, *the closure of* U_0 *being a* 3-*cell. Then there is a sequence of annuli,* $\{A_i\}$, *converging strongly to* A_0 *such that for each* i, A_i *lies in* M_i *and is bounded by simple closed curves* J_i' *and* J_i'' *such that* $J_i' \cup J_i'' = A_i \cap K_i$, *the sequences* $\{J_i'\}$ *and* $\{J_i''\}$ *converge* h-0-*regularly to* J_0' *and* J_0'' *and if* U_i *and* V_i *represent the components of* $M_i - A_i$, *the closure of* U_i *being a* 3-*cell,* $\{U_i\}$ *converges to* $A_0 \cup U_0$ *and* $\{V_i\}$ *converges to* $A_0 \cup V_0$.

Proof. It follows from the corollary to Theorem 2.14 that there are sequences $\{J_i'\}$ and $\{J_i''\}$ of simple closed curves converging h-0-regularly to J_0' and J_0'' such that for each i, J_i' and J_i'' lie in K_i. A slight modification of the proof of Lemma 2.13 implies the existence of a sequence $\{A_i'\}$ of singular annuli converging to A_0 such that for each i, A_i' lies in M_i and is bounded by $J_i' \cup J_i''$, $A_i' \cap K_i = J_i' \cup J_i''$, and there are no singularities on $J_i' \cup J_i''$. It follows from the generalization of Dehn's Lemma to annuli and other surface of genus 0 [15] that there is, arbitrarily close to A_i', a non-singular annulus A_i whose boundary is $J_i' \cup J_i''$ and which is such that

$A_i \cap K_i = J_i' \cup J_i''$. Each A_i may be so chosen that $\{A_i\}$ is the required sequence.

For each i, the set $K_i - (J_i' \cup J_i'')$ is the union of an open annulus B_i and two open discs D_i' and D_i'', $B_i \cup A_i$ is the boundary of the component V_i of $M_i - A_i$ and $D_i' \cup D_i'' \cup A_i$ is the boundary of the component U_i. Since $\{K_i\}$ converges 1-regularly to K_0, $\{A_i \cup B_i\}$ converges to $A_0 \cup B_0$ and $\{A_i \cup D_i' \cup D_i''\}$ converges to $A_0 \cup D_0' \cup D_0''$. The h-2-regular convergence of $\{M_i\}$ to M_0 implies that no point of $V_0 \cup B_0$ is a sequential limit point of any subsequence of $\{U_i\}$ and that no point of $U_0 - (D_0' \cup D_0'')$ is a sequential limit point of any subsequence of $\{V_i\}$. Thus $\{U_i\}$ and $\{V_i\}$ converge to $A_0 \cup U_0$ and $A_0 \cup V_0$.

COROLLARY. *If C_0 is a simple closed curve in A_0 separating J_0' from J_0'', then there is a sequence $\{C_i\}$ of simple closed curves converging strongly to C_0 such that for each i, $C_i \subset A_i$ and separates J_i' from J_i''.*

Proof. Since the sequence of 2-spheres, $\{A_i \cup D_i' \cup D_i''\}$ converges strongly to $A_0 \cup D_0' \cup D_0''$, the existence of a sequence $\{C_i\}$ converging strongly to C_0 such that for each i, $C_i \subset A_i$ is a direct consequence of an application of Lemma 2.5 to the natural mapping of a regular neighborhood in E^n of $A_0 \cup D_0' \cup D_0''$ onto $A_0 \cup D_0' \cup D''$. Suppose that U_0 is a regular neighborhood of C_0 consistently imbedded in W_0, a regular neighborhood of A_0, that p is the natural mapping, that p_i is a piecewise linear ϵ_i-approximation to $p \mid A_i$, and that γ_i is a non-trivial 1-cycle carried by C_i, which does not bound in $U_0 \cap M_i$. If C_i does not separate J_i' from J_i'' in A_i, then it bounds a disc in A_i. Thus γ_i bounds in A_i and $p_i(\gamma_i)$ bounds in A_0. Since $A_0 \cap U_0$ is a deformation retract of A_0, $p_i(\gamma_i)$ bounds in $A_0 \cap U_0$. Thus $g_i p_i(\gamma_i)$ bounds in $U_0 \cap M_i$. But, since for sufficiently large i and small ϵ_i, $g_i p_i(\gamma_i)$ is homologous to γ_i in $U_0 \cap M_i$, this is a contradiction.

LEMMA 3.4. *Suppose that A_0 is an annulus in M_0 whose boundary curves J_0' and J_0'' are such that $A_0 \cap K_0 = J_0' \cup J_0''$. Suppose, further, that B_0 is an annulus in $\operatorname{int} A_0$ whose boundary curves C_0' and C_0'' are such that C_0' separates J_0' from C_0'' which separates C_0' from J_0'' in A_0. Then there is a sequence $\{B_i\}$ of annuli converging strongly to B_0 such that for each i, $B_i \subset \operatorname{int} M_i$.*

Proof. Let $\{A_i\}$ be a sequence of annuli whose existence is implied by Lemma 3.3 such that (1) for each i, A_i is bounded by simple closed curves J_i' and J_i'' such that $A_i \cap K_i = J_i' \cup J_i''$, (2) the sequences $\{J_i'\}$ and $\{J_i''\}$ converge to J_0' and J_0'' h-0-regularly and (3) $\{A_i\}$ converges strongly to A_0.

Let ϵ be a positive number. Denote by Z_0' and Z_0'' simple closed curves in A_0 such that (1) Z_0' separates J_0' from C_0' and Z_0'' separates C_0'' from J_0'' and (2) the annuli in A_0, F_0' and F_0'', bounded by Z_0' and C_0' and C_0'' and Z_0'' respectively lie in $\epsilon/2$-neighborhoods of C_0' and C_0''. There are tori, T_0' and T_0'', which, together with their interiors, U_0' and U_0'', form regular $\epsilon/4$-neighborhoods of Z_0' and Z_0'', neither intersecting $B_0 \cup J_0' \cup J_0''$. From Lemma 2.16 it follows that there are sequences $\{T_i'\}$ and $\{T_i''\}$ of compact 2-manifolds converging strongly to T_0' and T_0'' such that for each i, T_i' and T_i'' are mutually exclusive subsets of $\operatorname{int} M_i$ with interiors U_i' and U_i'' and the sequences $\{U_i'\}$ and $\{U_i''\}$ converge to $U_0' \cup T_0'$ and $U_0'' \cup T_0''$. Denote by V_0' and V_0'' regular $\epsilon/4$-neighborhoods of Z_0' and Z_0'' in E^n such that $V_0'' \cap M_0$ is a regular neighborhood of $U_0'' \cup T_0''$ and $V_0' \cap M_0$ is a regular neighborhood of $U_0' \cup T_0'$ and by W_0 a regular $\epsilon/4$-neighborhood of A_0 in E^n with natural mapping p in which V_0' and V_0'' are consistently imbedded. For sufficiently large i, $U_i' \subset V_0'$ and $U_i'' \subset V_0''$.

It follows from the corollary to Lemma 3.3 that there are sequences $\{C_i'\}$ and $\{C_i''\}$ of simple closed curves converging strongly to C_0' and C_0'' such that for each i, each of C_i' and C_i'' lies in A_i and separates J_i' from J_i''. It is conceivable that C_i' fails to separate J_i' from C_i'' in A_i. If this is the case, C_i'' separates J_i' from C_i'. Denote by U_0 a regular $\epsilon/4$-neighborhood of C_0' not intersecting $J_0' \cup C_0''$, by T_0 its torus boundary, by $T_1, T_2, \cdots$ a sequence of 2-manifolds converging strongly to T_0 and by $U_1, U_2, \cdots$ the interiors in $M_1, M_2, \cdots$ of $T_1, T_2, \cdots$. The annuli $R_0, R_1, \cdots$ bounded by $J_0' \cup C_0', J_1' \cup C_0'', \cdots$ converge to a subset of A_0 which contains R_0. Hence for sufficiently large i, R_i intersects U_i and consequently T_i. Small changes may be made in T_i so that each component of $R_i \cap T_i$ is a simple closed curve. For sufficiently large i, $R_i \cap T_i$ separates J_i' from C_i'' in R_i and consequently one of the simple closed curve components, C_i, of $R_i \cap T_i$ separates J_i' from C_i''. If a non-trivial 1-cycle, γ_i, carried by C_i bounds in W_0, then, since C_i'' is deformable into C_i in R_i, C_i'' carries a non-trivial cycle which bounds in W_0, which contradicts the strong convergence of $\{C_i''\}$ to C_0''. If this argument is applied to a sequence of regular neighborhoods, U_0, of C_0' converging to C_0', a sequence $C_1, C_2, \cdots$ of simple closed curves is found which converges strongly to C_0' and which is such that for each i, C_i lies in A_i and separates J_i' from C_i''. Then C_i' may be replaced by C_i. Thus it may be assumed that C_i' does separate J_i' from C_i''. Denote the annulus in A_i bounded by $C_i' \cup C_i''$ by B_i'.

Let H_0', G_0 and H_0'' denote the three components of $W_0 - (V_0' \cup V_0'')$, H_0' containing J_0' and H_0'' containing J_0''. For sufficiently large i, $B_i' \subset W_0$.

$T_i' \cap B_i'$ separates $B_i' \cap G_0$ from $B_i' \cap H_0'$ and $B_i' \cap T_0''$ separates $B_i' \cap G_0$ from $B_i' \cap H_0''$. Small changes may be made in T_0' and T_i'' so that each component of $B_i' \cap T_i'$ and $B_i' \cap T_i''$ is a simple closed curve. If t is such a component and is not contractible in W_0, t separates C_i' from C_i'' in B_0', for otherwise, t would bound a disc in W_0, which is impossible. Thus the components of $B_i' \cap (T_i' \cup T_i'')$ which are not contractible in W_0 may be arranged in a sequence α, each element of which separates the elements preceding it from the elements following it in B_i'. Two elements of α will be called *joinable* if they are subsets of the same one of T_i' and T_i''. If t and t' are joinable in, say, T_i', then, since $p(t)$ is deformable into $p(t')$ in A_0 and consequently in $V_0' \cap A_0$, it follows from an application of Lemma 2.6 that for sufficiently large i, t is deformable into t' in $V_0' \cap M_i$ (i.e. $g_ip(t)$ is deformable into $g_ip(t')$ in $V_0' \cap M_i$ and for sufficiently large i and small ϵ_i, t is deformable into $g_ip(t)$ and t' into $g_ip(t')$ in $V_0' \cap M_i$).

Let s_1 denote the first element of α and s_2 the last element of α joinable to s_1. Let s_3 denote the first element of α between s_2 and C_2'', if such exists, not joinable to s_2 (otherwise, let $s_3 = C_2''$) and s_4 the last such element. Denote by R_1, R_2, and R_3 the annuli in A_1 bounded by C_i' and s_1, s_2 and s_3, and s_4 and C_i''. If t is a component of $R_j \cap (T_i' \cup T_i'')$ other than s_1, s_2, s_3, or s_4, t is not in α, so bounds a singular disc D_t in $V_0' \cap M_i$ or $V_0'' \cap M_i$ and a non-singular disc in A_i. If t is not contained in any other such disc in A_i, replace the disc in A_i bounded by t by D_t for each such t and replace the annuli in A_i bounded by s_1 and s_2 and s_3 and s_4 by singular annuli with the same boundaries in $V_0' \cap M_i$ and $V_0'' \cap M_i$. In this way, B_i' is replaced by a singular annulus having no singularities on its boundary, $C_i' \cup C_i''$, and not intersecting $H_0' \cup H_0''$. Thus it follows from the extension of Dehn's Lemma that $C_i' \cup C_i''$ bounds a non-singular annulus B_i in $M_i \cap W_0$ which does not intersect $H_0' \cup H_0''$ and thus lies in an ϵ-neighborhood of B_0. Since ϵ was arbitrary, the existence of the required sequence is established.

COROLLARY. *The Lemma remains true if one of the curves C_0' and C_0'' is J_0' or J_0''.*

Note. The extension of Dehn's Lemma demonstrates that, since B_i' lies in the annulus A_i, B_i may be constructed so that $\operatorname{int} B_i \cap (A_i - B_i')$ lies outside some small neighborhood of $C_i' \cup C_i''$.

LEMMA 3.5. *Suppose that A_0 is an annulus in M_0 bounded by simple closed surves J_0' and J_0'' such that $A_0 \cap K_0 = J_0' \cup J_0''$. Suppose, further, that $J_0' = t_{00}, t_{01}, \cdots, t_{0m} = J_0''$ is a sequence of simple closed curves in A_0,*

in that order, each separating J_0' from J_0'', and that $A_{01}, A_{02}, \cdots, A_{0m}$ are the annuli in A_0 bounded by consecutive pairs of elements of $\{t_{0j}\}$. Then there is a sequence $\{A_i\}$ of annuli converging strongly to A_0 such that for each i, (1) A_i lies in M_i, is bounded by simple closed curves J_i' and J_i'' and $A_i \cap K_i = J_i' \cup J_i''$, (2) there is a sequence $J_i' = t_{i0}, \cdots, t_{im} = J_i''$ of simple closed curves in A_i in that order, A_{ij} denoting the annulus in A_i whose boundary is $t_{ij-1} \cup t_{ij}$ and (3) the elements of $\{t_{ij}\}_j$ may be so selected that $\{t_{ij}\}_i$ converges strongly to t_{0j} and $\{A_{ij}\}_i$ converges to A_{0j}.

Proof. Suppose that ϵ is a positive number, W_0 is a regular ϵ-neighborhood in E^n of A_0 and for each j, W_{0j} is a regular ϵ-neighborhood in E^n of A_{0j} consistently imbedded in W_0 such that $W_{0j-1} \cap W_{0j} = V_{0j-1}$ is a regular neighborhood of t_{0j-1}, $W_{0j} \cap W_{0k} = 0$ unless $|j-k| \leqq 1$, and $W_{0j} \cap K_0 = 0$ unless $j = 0, m$. If, for sufficiently large i, an annulus A_i may be found which satisfies conditions (1) and (2) of the statement of the lemma and which is such that $A_{ij} \subset W_{0j}$ and $t_{ij} \subset V_{0j}$ and is not contractible in V_{0j}, then the lemma is proved. (If t_{ij} carries a nontrivial cycle which bounds in V_{0j}, then $p_{ij}(t_{ij})$ carries a nontrivial cycle which bounds in $V_{0j} \cap A_0$, where p_{ij} ϵ_i-approximates the projection map of V_{0j} into $V_{0j} \cap A_0$. Thus $p_{ij}(t_{ij})$ is contractible in $V_{0j} \cap A_0$ and t_{0j} is contractible in V_{0j}.)

For each $j = 2, 3, \cdots, m$, let s_{0j} denote a simple closed curve in $A_{0j} \cap V_{0j-1}$ and let s_{01} be a simple closed curve in an ϵ-neighborhood of t_{00} such that for each j, s_{0j} separates t_{0j-1} from t_{0j} in A_0. Let R_{0j} and T_{0j} denote the annuli in A_{0j} bounded by t_{0j-1} and s_{0j} and s_{0j} and t_{0j}. Lemma 3.4, particularly the remarks in the second part of its proof, implies the existence, for sufficiently large i, of sequences of simple closed curves, $t_{i0}', t_{i1}', \cdots, t_{im}'$ and $s_{i1}', s_{i2}', \cdots, s_{im}'$ and sequences of mutually exclusive annuli $\{R_{ij}'\}_j$ and $\{T_{ij}'\}_j$ such that for each j, (1) R_{ij}' is bounded by $t_{ij-1}' \cup s_{ij}'$ and T_{ij}' is bounded by $s_{ij}' \cup t_{ij}'$, (2) $R_{ij}' \subset V_{ij-1}$, (3) $T_{ij}' \subset W_{0j}$, (4) s_{ij}' and t_{ij-1}' are not contractible in V_{0j-1}, and (5) $(R_{i1}' \cup T_{i1}') \cap K_i = t_{i0}'$ and $(R_{im}' \cup T_{im}') \cap K_i = t_{im}'$. Small changes may be made in these annuli so that each component of $(T_{ij-1}' \cup T_{ij}') \cap R_{ij}'$ is a simple closed curve. (See the note following the proof of Lemma 3.4.) Arrange the components of $T_{ij}' \cap (R_{ij}' \cup R_{ij+1}')$ which are not contractible in W_0 in a sequence α_{ij}, the order in the secquence being determined by the order of the components in T_{ij}' from s_{ij}' to t_{ij}'.

Denote t_{i0}' by t_{i0}, t_{im}' by t_{im}, the last element of α_{ij} in $R_{ij}' \cap T_{ij}'$ by s_{ij} and the first element of α_{ij} following s_{ij} by t_{ij}. The simple closed curve t_{ij} lies in R_{ij+1}'. Let R_{ij}'' denote the annulus in R_{ij}' bounded by t_{ij-1} and s_{ij} and T_{ij}'' the annulus in T_{ij}' bounded by s_{ij} and t_{ij}. Then $T_{ij}'' \cap T_{ik}'' = 0$ and

$R_{ij}'' \cap R_{ik}'' = 0$ unless $j = k$ and each component of $(R_{ij}'' \cup R_{ij+1}'') \cap T_{ij}''$ other than s_{ij} and t_{ij} is a simple closed curve which is contractible in W_{0j} and hence bounds a disc in R_{ij}'' or R_{ij+1}''.

Consider $H_{ij} = R_{ij}'' \cap (T_{ij-1}'' \cup T_{ij}'')$. If t is a simple closed curve component of H_{ij} whose interior, D, in R_{ij}'' does not intersect H_{ij} and t lies in, say, T_{ij-1}'', t bounds a disc F in T_{ij-1}''. Replace F by D and move the adjusted T_{ij-1}'' slightly away from R_{ij}'' in such a way that no new intersections with any T_{ij}'' or R_{ij}'' are added. The adjusted T_{ij-1}'' still lies in W_{0j-1} and has one less simple closed curve in common with R_{ij}''. If this process is repeated until all the components of H_{ij} are removed except t_{ij-1} and s_{ij}, T_{ij}'' and R_{ij}'' are replaced by annuli T_{ij} and R_{ij} such that (1) $T_{ij} \cap T_{ih} = 0$ unless $j = h$ and $R_{ij} \cap R_{ih} = 0$ unless $j = h$, (2) $T_{ij} \cap R_{ij} = s_{ij}$, (3) $T_{ij} \cap R_{ij+1} = t_{ij}$, (4) $R_{ij} \subset V_{0j}$, and (5) $T_{ij} \subset W_{0j}$. The set $R_{ij} \cup T_{ij}$ is thus an annulus A_{ij} in W_{0j} and the annulus $A_i = \cup_j A_{ij}$, together with the sequences $\{A_{ij}\}_j$ and $\{t_{ij}\}_j$, satisfies conditions (1) and (2) of the statement of the lemma and the conditions stated in the first paragraph of this proof. This implies the truth of Lemma 3. 5.

LEMMA 3. 6. *Suppose that R is a 3-cell with boundary S, t is an unknotted (polyhedral) arc in R with endpoints P' and P'' such that $t \cap S = P' \cup P''$ and A is a (polyhedral) annulus in R with boundary curves J' and J'' such that* (1) *$A \cap S = J' \cup J''$,* (2) *the discs D' and D'' in S bounded by J' and J'' contain P' and P'' respectively and* (3) *t lies in the component of $R - A$ whose closure is a 3-cell. Then A is unknotted.*

Proof. There is a disc D in R with boundary J such that $D \cap S = J$, $J \cap J'$ and $J \cap J''$ each consists of just two points and $t \subset D$. Small changes may be made in D so that each component of $D \cap A$ is either an arc or a simple closed curve. If a component s of $D \cap A$ is an arc, the endpoints of s lie in $J' \cup J''$. If both endpoints lie in, say, J', then $s \cup (D' \cap J)$ is a simple closed curve in D, so s intersects t—a contradiction. Hence one endpoint of s lies in J' and the other in J''. There is only one other arc as a component of $D \cap A$. Call it u. If a simple closed curve, w, is a component of $D \cap A$, it bounds a disc, F, in A, for if this is false, w and J' bound an annulus, K, in A and w bounds a disc, E, in D which does not intersect t. Thus $E \cup K$ is a singular disc in $R - t$ bounded by J'. This, however, is impossible. If $F - w$ does not intersect $D \cap A$, replace E in D by F and move it slightly away from A so that the adjusted D has one less simple closed curve in common with A. Repeat this process until a disc D' is obtained which contains t, is bounded by J and is such that $D \cap A = s \cup u$. That A is unknotted now follows readily.

THEOREM 3.7. *If t_0 is an unknotted arc in M_0 with endpoints P_0 and Q_0 such that $t_0 \cap K_0 = P_0 \cup Q_0$, then there is a sequence of arcs $\{t_i\}$ with endpoints $\{P_i \cup Q_i\}$ converging h-0-regularly to t_0 such that for each i, (1) t_i lies in M_i and is unknotted in M_i, (2) $t_i \cap K_i = P_i \cup Q_i$ and (3) the sequences $\{P_i\}$ and $\{Q_i\}$ converge to P_0 and Q_0.*

Proof. Let $P_0 = P_{00}, P_{01}, \cdots, P_{0m} = Q_0$ be a sequence of points of t_0 in that order, let W_0 be a regular neighborhood in E^n of t_0 and for each j, let W_{0j} be a regular neighborhood of the subarc $P_{0j-1}P_{0j}$ of t_0 which is consistently imbedded in W_0 and is such that $W_{0j} \cap W_{0j+1}$ is a regular neighborhood of P_{0j}, $W_{0j} \cap W_{0k} = 0$ unless $|j-k| \leqq 1$, and $W_{0j} \cap K_0 = 0$ unless $j = 0, m$. The theorem will be proved if it can be shown that for sufficiently large i there is an arc t_i satisfying conditions (1) and (2) which lies in W_0 and is such that for each j, $t_i \cap (W_{0j} \cup W_{0j+1})$ has only one component which intersects both W_{0j-1} and W_{0j+2}.

There is an unknotted annulus A_0 in $W_0 \cap M_0$ with boundary $t_{00} \cup t_{0m}$ such that (1) $A_0 \cap K_0 = t_{00} \cup t_{0m}$, (2) t lies in the component of $M_0 - A_0$ whose closure is a 3-cell and (3) there are simple closed curves $t_{00}, t_{01}, \cdots, t_{0m}$ in that order in A_0 such that for each j, $t_{0j-1} \cup t_{0j}$ bounds an annulus A_{0j} in $A_0 \cap W_{0j}$. From Lemma 3.2 it follows that there is a sequence $s_1, s_2, \cdots$ of arcs converging to t_0 such that for each i, s_i is unknotted in M_i, s_i has endpoints P_i' and Q_i', $s_i \cap K_i = P_i' \cup Q_i'$ and the sequences $\{P_i'\}$ and $\{Q_i'\}$ converge to P_0 and Q_0. It follows from Lemmas 3.3 and 3.5 that for sufficiently large i there is an annulus A_i in $W_0 \cap M_i$ bounded by simple closed curves t_{i0} and t_{im} such that (1) $A_i \cap K_i = t_{i0} \cup t_{im}$, (2) s_i lies in the component of $M_i - A_i$ whose closure is a 3-cell and (3) there are simple closed curves $t_{i0}, \cdots, t_{im}$ in that order in A_i such that for each j, $t_{ij-1} \cup t_{ij}$ bounds an annulus A_{ij} in $A_i \cap W_{0j}$. Let t_i denote an arc in A_i with endpoints P_i and Q_i such that (1) $t_i \cap t_{i0} = P_i$, (2) $t_i \cap t_{im} = Q_i$, (3) for each j, $t_i \cap A_{ij}$ is an arc. Since s_i is unknotted, it follows from Lemma 3.6 that A_i and therefore t_i is unknotted. Clearly there are not two components of $t_i \cap (W_{0j} \cup W_{0j+1})$ which intersect both W_{0j-1} and W_{0j+2} so that t_i is the required arc.

LEMMA 3.8. *Let t_0' and t_0'' be mutually exclusive unknotted arcs in M_0 such that $t_0' \cap K_0$ and $t_0'' \cap K_0$ are the unions of the endpoints of t_0' and t_0'' and let D_0 be a disc in M_0 whose boundary, J_0, contains t_0' and t_0'' and is such that $\mathrm{cl}(J_0 - (t_0' \cup t_0'')) = D_0 \cap K_0$. Then if $\{t_i'\}$ and $\{t_i''\}$ are sequences of arcs converging regularly to t_0' and t_0'' such that for each i, t_i' and t_i'' are unknotted in M_i and meet K_i only in their endpoints, there exists*

a sequence of discs $D_1, D_2, \cdots$ *converging to* D_0 *whose boundaries* $J_1, J_2, \cdots$ *converge regularly to* J_0 *such that for each* i, $t_i' \cup t_i'' \subset J_i$ *and*

$$\operatorname{cl}(J_i - (t_i' \cup t_i'')) = D_i \cap K_i.$$

Proof. There is a disc E_0 in M_0 which contains D_0 and whose boundary, B_0 is such that $E_0 \cap K_0 = B_0$. Also, there is a disc F_0 in M_0 with boundary C_0 such that $F_0 \cap E_0$ is an arc t_0 which separates t_0' from t_0'' in E_0, $F_0 \cap K_0 = C_0$, and $C_0 \cap B_0$ is the union of the endpoints of t_0. From Lemma 2.13 it follows that there is a sequence of discs, $F_1, F_2, \cdots$ converging to F_0 whose boundaries $C_1, C_2, \cdots$ converge regularly to C_0 and are such that for each i, $F_i \cap K_i = C_i$. If, for each i, U_i' and U_i'' denote the closures of the components of $M_i - F_i$, where $t_0' \subset U_0'$ and $t_0'' \subset U_0''$, the h-2-regularity of the convergence of $\{M_i\}$ to M_0 implies that the sequences $\{U_i'\}$ and $\{U_i''\}$ converge to U_0' and U_0''. Thus, for sufficiently large i, $t_i' \subset U_i'$ and $t_i'' \subset U_i''$. It is clear, then, that there is a disc E_i''' in M_i containing $t_i' \cup t_i''$ whose boundary, B_i, is such that $E_i''' \cap K_i = B_i$ and the sequence $\{B_i\}$ may be so selected that it converges regularly to B_0.

The E_i''' must be adjusted in such a way that the resulting sequence converges to E_0. Let V_0 be a regular ϵ-neighborhood of E_0 in M_0 whose boundary consists of an annulus in K_0 and the discs E_0' and E_0'' whose boundaries, B_0' and B_0'' are such that $E_0' \cap K_0 = B_0'$ and $E_0'' \cap K_0 = B_0''$. There are sequences $\{E_i'\}$ and $\{E_i''\}$ of discs converging strongly to E_0' and E_0'' such that for each i, $E_i' \cup E_i'' \subset M_i$, $E_i' \cap K_i = \operatorname{bdry} E_i'$ and $E_i'' \cap K_i = \operatorname{bdry} E_i''$. Denote by V_i the closure of the component of $M_i - (E_i' \cup E_i'')$ which contains $t_i' \cup t_i''$. Clearly $B_i \subset V_i$ and $\{V_i\}$ converges to V_0. The E_i' and E_i'' may be adjusted so that each component of $(E_i' \cup E_i'') \cap E_i'''$ is a simple closed curve. A push-pull argument proves that each E_i''' may be replaced by a disc E_i which contains $t_i' \cup t_i''$, lies in V_i, is bounded by B_i and meets K_i only in B_i. Since ϵ was arbitrary, the existence of a sequence of discs $\{E_i\}$ converging to E_0 such that for each i, $E_i \subset M_i$, $E_i \cap K_i = B_i$ and $t_i' \cup t_i'' \subset E_i$ is established. The proof of Lemma 3.4 may now be applied, with the obvious changes, to construct the required sequence $\{D_i\}$.

Note. In the application of Dehn's Lemma here, D_i may be so constructed that $\operatorname{int} D_i \cap (E_i - D_i')$, where D_i' is the disc in E_i replaced by D_i, does not intersect some small neighborhood of $t_i' \cup t_i''$. Also, the above proof demonstrates that if G_0 is the closure of the component of $E_0 - D_0$ whose boundary contains t_0' and $G_1, G_2, \cdots$ is a sequence of discs converging to G_0 such that for each i, $G_i \subset M_i$ and $\operatorname{bdry} G_i = t_i' \cup (G_i \cap K_i)$, then E_i may be

so constructed that $G_i \subset E_i$. Thus $(\operatorname{int} D_i) \cap G_i$ has no points in some small neighborhood of t_i'.

LEMMA 3.9. *Suppose that D_0 is a disc in M_0 with boundary J_0 such that $D_0 \cap K_0 = J_0$. Suppose, further, that $t_{00}, \cdots, t_{0m}$ is a sequence of arcs in D_0 such that (1) $t_{00} \cup t_{0m} \subset J_0$, (2) $t_{0j} \cap J_0$ is, for $j \neq 0, m$, the union of the endpoints of t_{0j} and (3) t_{0j} separates t_{0j-1} from t_{0j+1} in D_0. Denote by D_0 the disc in D_{0j} which is the closure of the component of $D_0 - \cup t_{0j}$ whose boundary contains $t_{0j-1} \cup t_{0j}$. Then there is a sequence $\{D_i\}$ of discs converging to D_0 whose boundaries $\{J_i\}$ converge regularly to J_0 such that for each i, (1) $D_i \cap K_i = J_i$ and (2) there are arcs $t_{i0}, t_{i1}, \cdots, t_{im}$ in D_i such that $t_{i0} \cup t_{im} \subset J_i$, $t_{ij} \cap J_i$ is the union of the endpoints of t_{ij} for $j \neq 0, m$ and $t_{ij-1} \cup t_{ij}$ lies on the boundary of a disc D_{ij}, which is the closure of a component of $D_i - \cup_j t_{ij}$. Furthermore, the sequence $\{t_{ij}\}_i$ converges h-0-regularly to t_{0j} and $\{D_{ij}\}_i$ converges to D_{0j}.*

Proof. Let W_0 be a regular neighborhood of D_0 in E^n and for each j, let W_{0j} be a regular neighborhood of D_{0j} consistently imbedded in W_0 such that $W_{0j} \cap W_{0j+1} = V_{0j}$ is a regular neighborhood of t_{0j} consistently imbedded in W_0, $W_{0j} \cap W_{0k} = 0$ unless $|j-k| \leqq 1$, and $W_{0j} \cap K_0 = 0$ unless $j = 0, m$. Let V_{00} and V_{0m} be regular neighborhoods of t_{00} and t_{0m} consistently imbedded in W_0 and meeting no other V_{0j}. If for each positive number ϵ it can be shown that for sufficiently large i, there is a disc D_i satisfying conditions (1) and (2) such that for each j, $D_{ij} \subset W_{ij}$ and t_{ij} is ϵ-homeomorphic to t_{0j}, then the lemma is proved.

For each j, let s_{0j} be an arc in $D_{0j} \cap V_{0j-1}$ which lies except for its endpoints in $\operatorname{int} D_{0j}$ and separates t_{0j-1} from t_{0j} and let R_{0j} and T_{0j} be the discs into which s_{0j} separates D_{0j}, R_{0j} containing t_{0j-1} and T_{0j} containing t_{0j}. Construct s_{0j} so that R_{0j} lies in V_{0j-1}. Theorem 3.7 and Lemma 3.8 imply that for sufficiently large i, there are, for each j, (1) a simple closed curve J_i in K_i which is ϵ-homeomorphic to J_0, (2) arcs t_{ij-1} and s_{ij} in V_{0j-1} which are ϵ-homeomorphic to t_{0j-1} and s_{0j}, t_{i0} lying in J_i, t_{ij-1} lying except for its endpoints in $\operatorname{int} M_i$ if $j > 1$, (3) an arc t_{im} in J_i which is ϵ-homeomorphic to t_{0m}, (4) a disc R_{ij} which lies in V_{0j-1}, is bounded by $t_{ij-1} \cup s_{ij}$ and portions of J_i and meets K_i only in these portions of J_i and (5) a disc T_{ij}' which lies in W_{0j}, is bounded by $s_{ij} \cup t_{ij}$ and portions of J_i and meets K_i only in these portions of J_i. Clearly these arcs and discs may be so chosen that no two arcs intersect and $T_{ij}' \cap T_{ik}' = 0$ unless $j = k$.

Small changes in the R_{ij} may be made so that each component of $H_{ij} = (T_{ij-1}' \cup T_{ij}') \cap R_{ij}$ except t_{ij-1} and s_{ij} is a simple closed curve. (See

the note following Lemma 3.8.) If t is such a component in, say, T_{ij-1}', whose interior, E, in R_{ij} does not intersect H_{ij}, replace its interior in T_{ij-1}' by E and move it slightly away from R_{ij} so that H_{ij} has one less component and the adjusted T_{ij-1}', which still lies in W_{0j-1}, has no new intersection with any R_{ij} or T_{ij}'. Repeat this process until each T_{ij}' is replaced by a disc T_{ij} such that $T_{ij} \cap R_{ij} = s_{ij}$, $T_{ij-1} \cap R_{ij} = t_{ij-1}$ and $\cup (T_{ij} \cup R_{ij})$ is a disc D_i such that $D_i \cap K_i = J_i$. Denote $R_{ij} \cup T_{ij}$ by D_{ij}. The disc D_{ij} lies in W_{ij}, t_{ij} is ϵ-homeomorphic to t_{0j} and D_i satisfies conditions (1) and (2) of the statement of the lemma. Thus the existence of the required sequence is proved.

THEOREM 3.10. *If D_0 is a disc in M_0 with boundary J_0 such that $D_0 \cap K_0 = J_0$, then there is a sequence of discs $\{D_i\}$ with boundaries $\{J_i\}$ converging h-0-regularly to D_0 such that for each i, $D_i \subset M_i$ and $D_i \cap K_i = J_i$.*

Proof. Let $t_{00}, t_{01}, \cdots, t_{0m}$ be a sequence of arcs in D_0 such that $t_{00} \cup t_{0m} \subset J_0$, $t_{0j} \cap J_0$, for $j \neq 0, m$, is the union of the endpoint of t_{0j}, t_{0j} separates t_{0j-1} from t_{0j+1} in D_0 and $t_{0j-1} \cup t_{0j}$ is on the boundary of a disc E_{ij} which is the closure of a component of $D_0 - \cup t_{0j}$. Let $s_{00}, s_{01}, \cdots, s_{0m}$ be a sequence of arcs in D_0 such that $s_{00} \cup s_{0m} \subset J_0$, $s_{0j} \cap J_0$, for $j \neq 0, m$, is the union of the endpoints of s_{0j}, s_{0j} separates s_{0j-1} from s_{0j+1} in D_0 and $s_{0j} \cap t_{0k}$ is a point, P_{0jk}. Denote by D_{0j} the disc which is the closure of the component of $D_0 - \cup s_{0j}$ bounded in part by $s_{0j-1} \cup s_{0j}$, and by F_{0jk} the disc $D_{0j} \cap E_{0k}$. Let W be a regular neighborhood of D_0 in E^n and U_j a regular neighborhood of D_{0j} consistently imbedded in W such that $U_j \cap U_{j+1}$ is a regular neighborhood of s_{0j} consistently imbedded in W and $U_j \cup K_0 = 0$ unless $j = 0, m$. Suppose that $U_j \cap U_{j'} = 0$ unless $|j - j'| \leqq 1$. Also, let V_k be a regular neighborhood of E_{0k} consistently imbedded in W such that $V_k \cap V_{k+1}$ is a regular neighborhood of t_{0k} consistently imbedded in W, $V_k \cap V_{k'} = 0$ unless $|k - k'| \leqq 1$, $V_{0k} \cap U_{0j}$ is a regular neighborhood W_{jk} of F_{0jk} consistently imbedded in W, and $V_k \cap K_0 = 0$ unless $k = 0, m$.

It follows from the Lemma 3.9 that there is a sequence $\{D_i'\}$ of discs converging to D_0 whose boundaries, $\{J_i\}$ converge 0-regularly to J_0 such that for each i, (1) $D_i' \cap K_i = J_i$ and (2) there are arcs $t_{i0}, t_{i1}, \cdots, t_{im}$ in D_i' such that $t_{i0} \cup t_{im} \subset J_i$, $t_{ik} \cap J_i$ is, for $k \neq 0, m$, the union of the endpoints of t_{ik}, and $t_{ik-1} \cup t_{ik}$ lies in the boundary of a disc, E_{ik}', which is the closure of a component of $D_i' - \cup_k t_{ik}$ and lies, for sufficiently large i, in V_k. Furthermore, for each k, the sequence $\{t_{ik}\}_i$ converges regularly to t_{0k} and $\{E_{ik}'\}$ converges to E_{0k}. An application of the proof of Lemma 3.2 demonstrates the existence, for sufficiently large i and each j, of an arc s_{ij} in $U_j \cap U_{j+1} \cap D_i$ such that $s_{i0} \cup s_{im} \subset J_i$, $s_{ij} \cap J_i$ is the union of the endpoints of s_{ij} for $j \neq 0, m$

and $s_{ij} \cap t_{ik}$ is a point, P_{ijk}. Denote by D_{ij}' the closure of the component of $D_i' - \bigcup_j s_{ij}$ whose boundary contains $s_{ij-1} \cup s_{ij}$ and by F_{ijk}' the disc $D_{ij}' \cap E_{ik}'$.

It may be assumed that each W_{jk} has diameter less than $\epsilon/100$. The theorem will be proved if it can be shown that there is a positive number δ such that for sufficiently large i, there is a disc D_i in $W \cap M_i$ such that $D_i \cap K_i = J_i$ and each pair of points in D_i whose distance apart is less than δ bounds an arc in D_i of diameter less than ϵ. Let δ be such that if p and q are points in W whose distance apart is less than δ, then for some j and k,

$$p \cup q \subset W_{jk} - W_{jk} \cap (V_{k'} \cup U_{j'}),$$

where $|k - k'| = |j - j'| = 1$. The remainder of the proof is devoted to the demonstration of the existence for sufficiently large i of a disc D_i satisfying the above conditions with respect to δ.

For each k, let V_k' denote a regular neighborhood of E_{0k} such that $V_k' \subset \operatorname{int} V_k$. Denote $V_k' \cap U_j$ by W_{jk}'. It may be assumed that for sufficiently large i, $E_{ij}' \subset V_j'$. Denote by A_{0j} an annulus in $U_j \cap U_{j+1} \cap M_0$ such that (1) $A_{0j} \cap K_0 = \operatorname{bdry} A_{0j}$, (2) s_{0j} lies in the component of $M_0 - A_{0j}$ whose closure is a 3-cell and (3) $A_{0j} \cap W_{jk}'$ and $A_{0j} \cap W_{j+1k}'$ are annuli. It follows from Lemma 3.5 that for sufficiently large i there is an annulus A_{ij} in $U_j \cap U_{j+1} \cap M_i$ such that (1) $A_{ij} \cap K_i = \operatorname{bdry} A_{ij}$, (2) s_{ij} lies in the component of $M_i - A_{ij}$ whose closure is a 3-cell, (3) each simple closed curve in $A_{ij} \cap W_{jk}'$ which is contractible in A_{ij} bounds a disc in $A_{ij} \cap W_{jk}$, and (4) $A_{ij} \cap D_i'$ separates $D_i' \cap \bigcup U_{j'}$, $j' < j$, from $D_i' \cap \bigcup U_{j'}$, $j' > j + 1$. Small adjustments may be made in A_{ij} so that each component of $A_{ij} \cap D_{ij}'$, except for an arc in D_{ij}' from t_{00} to t_{0m} and one in D_{ij+1}', is a simple closed curve which is contractible in A_{ij}. If $j' \neq j$, $j+1$, each component of $D_{ij'}' \cap A_{ij}$ is, for some k, a subset of W_{jk}'.

Suppose that t is a component of $A_{ij} \cap D_i'$ which for some k lies in W_{jk}' and suppose, further, that the interior, E, of t in A_{ij} does not intersect D_i'. The interior, F, of t in D_i' lies either in $F_{ij'k'}'$, for some $j' \neq j$, $j+1$, $|k - k'| \leqq 1$, or in $F_{ijk-1}' \cup F_{ijk}' \cup F_{ijk+1}'$ or in $F_{ij+1k-1}' \cup F_{ij+1k}' \cup F_{ij+1k+1}'$ and may be replaced by E and then moved slightly away from A_{ij} so that the number of components of $D_i' \cap A_{ij}$ is reduced. If $\operatorname{int} E$ does intersect D_i', consider a component t' of $E \cap D_i'$ whise interior, E', does not intersect D_i'. Let F' denote the interior of t' in D_i'. The set E' lies in W_{jk} and F' can be replaced by E' as above. If this process is repeated, each $F_{ij'k}$, $j' \neq j$, $j+1$, is replaced by a disc $F_{ij'k}''$ whose boundary is that of $F_{ij'k}'$, which does not intersect both components of $W - (U_j \cap U_{j+1})$—i. e. does not intersect A_{ij}—and lies in V_k. Also, certain discs in D_{ij}' and D_{ij+1}' whose boundaries lie in

some W_{jk}' (or, in some cases, in W_{jk}) are replaced by discs in W_{jk}. Furthermore, $\cup F_{ijk}''$ is a disc, no s_{ij} is changed and no new intersections with any A_{ij} are added. This process is applied to each A_{ij}, starting with A_{i0}. Consider F_{ijk}'. After this process has been applied to each $A_{ij'}$, $j' < j-1$, F_{ijk}' has been replaced by a disc F_{ijk}'' whose boundary is that of F_{ijk}', which lies in V_k and $\cup U_{j'}$, $j' \geqq j-1$, and whose intersection with $A_{ij'}$, $j' \geqq j-1$, still lies in $W_{j'k}'$. After this process is applied to A_{ij-1} and A_{ij}, F_{ijk}'' is replaced by a disc F_{ijk} which lies in $W_{jk-1} \cup W_{jk} \cup W_{jk+1}$. The disc F_{ijk} is not affected by the action on $A_{ij'}$ for $j' > j$. Also, the resulting $\cup F_{ijk}$ is a disc D_i whose boundary is J_i.

Suppose that p and q are points of D_i whose distance apart is les than δ and that $p \cup q \subset W_{jk} - W_{jk} \cap (V_{k-1} \cup U_{j-1})$. Then for sufficiently large i, $p \cup q$ is a subset of the union of F_{ijk}, F_{ijk+1}, F_{ijk+2}, F_{ijk-1}, F_{ij+1k}, $F_{ij+1k+1}$, $F_{ij+1k+2}$, and $F_{ij+1k-1}$ which, since each W_{jk} has diameter less than $\epsilon/100$, certainly has diameter less than ϵ. This completes the proof of the theorem.

LEMMA 3.11. *If R_0 is a 3-cell in M_0 bounded by the 2-sphere S_0 and $S_1, S_2, \cdots$ is a sequence of 2-spheres converging h-0-regularly to S_0 such that for each i, S_i bounds the 3-cell R_i in M_i, then the sequence $\{R_i\}$ converges h-2-regularly to R_0.*

Proof. The h-2-regularity of the convergence of $\{M_i\}$ to M_0 implies that the convergence of $\{R_i\}$ to R_0 is regular at each point of $\operatorname{int} R_0$. Suppose that P is a point of S_0 and ϵ is a positive number. There is a positive number $\delta' < \epsilon/2$ such that each singular 2-sphere in M_i of diameter less than $2\delta'$ bounds a singular 3-cell in M_i of diameter less than $\epsilon/2$ and there is a positive number $\delta < \delta'/2$ such that each singular 1-sphere in S_i of diameter less than δ bounds a singular 2-cell in S_i of diameter less than $\delta'/2$.

Suppose that J_i is a singular j-sphere, $j \leqq 2$, which may be assumed to be polyhedral, in $R_i \cap S(P, \delta/4)$. It may be assumed that $J \cap S_i = 0$. Denote by S_0' and R_0' the boundary and closure respectively of the spherical neighborhood $M_0 \cap S(P, \delta/2)$. There is a sequence $\{S_i'\}$ of 2-spheres converging strongly to S_0' such that for each i, S_i' bounds a 3-cell, R_i', in M_i. The sequences $\{R_i'\}$ and $\{M_i - R_i'\}$ converge to R_0' and $\operatorname{cl}(M_0 - R_0')$. Hence, for sufficiently large i, $J_i \subset \operatorname{int} R_i'$. Adjust each S_i' slightly so that each component, K, of $S_i' \cap R_i$ is a subset of a disc which is bounded by a finite number of simple closed curves, each of which bounds a (non-singular) disc in S_i of diameter less than $\delta'/2$. Add these discs to K, moving them slightly away from S_i in such a way that K becomes a non-singular 2-sphere, K', in $S(P, \delta') \cap R_i$. At least one such 2-sphere, call it K_i, has J_i in its interior.

Since K_i has diameter less than $2\delta'$, it bounds a 3-cell C_i in M_i, consequently in R_i, of diameter less than $\epsilon/2$. But C_i contains J_i and therefore J_i bounds a $(j+1)$-cell in C_i which is a subset of $R_i \cap S(P,\epsilon)$. This proves the lemma.

THEOREM 3.12. *The sequence $M_1, M_2, \cdots$ converges to M_0 completely regularly.*

Proof. Suppose that ϵ is a positive number. It will be shown that for sufficiently large i, there is an ϵ-homeomorphism of M_0 onto M_i. The proof will consist of the construction for each i of a certain subdivision of M_i similar to that obtained for a geometric cube in E^3 by sections of planes parallel to its faces.

Denote by $D_{00}, D_{01}, \cdots, D_{0m}, E_{00}, E_{01}, \cdots, E_{0m}, F_{00}, F_{01}, \cdots, F_{0m}$ three sequences of mutually exclusive discs in M_0 such that (1) $D_{00} \cup D_{0m} \cup E_{00} \cup E_{0m} \cup F_{00} \cup F_{0m} \subset K_0$, (2) for each $p \neq 0$, (i) $D_{0p} \cap K_0 = \text{bdry}\, D_{0p}$, $E_{0p} \cap K_0 = \text{bdry}\, E_{0p}$ and $F_{0p} \cap K_0 = \text{bdry}\, F_{0p}$ and (ii) D_{0p} separates D_{0p-1} from D_{0p+1}, E_{0p} separates E_{0p-1} from E_{0p+1} and F_{0p} separates F_{0p-1} from F_{0p+1} in M_0, (3) for each p, q, r, $D_{0p} \cap E_{0q}$, $E_{0q} \cap F_{0r}$, and $D_{0p} \cap F_{0r}$ are arcs and $D_{0p} \cap E_{0q} \cap F_{0r}$ is a point and (4) each component of $M - \cup (D_{0p} \cup E_{0p} \cup F_{0p})$ has diameter less than $\epsilon/2$.

It follows from Theorem 3.10 that there is, for each p, a sequence $\{D_{ip}\}$ of discs converging h-0-regularly to D_{0p} such that for each i, (1) D_{ip} is in M_i and $D_{ip} \cap K_i = \text{bdry}\, D_{ip}$, $(p \neq 0, m)$, (2) $D_{i0} \cup D_{im} \subset K_i$, (3) D_{ip} separates D_{ip-1} *from* D_{ip+1} in M_i for $p \neq 0, m$. Denote by R_{ip} the closure of the component of $M_i - \cup_p D_{ip}$ whose boundary contains $D_{ip-1} \cup D_{ip}$. It follows from Lemma 3.11 that for each p, the sequence of 3-cells, $\{R_{ip}\}_i$ converges h-2-regularly to R_{0p} and that the lemmas and theorems already proved can be applied to these sequences.

If Theorem 3.10 is applied to each such sequence, it is shown that for each q there is a sequence $\{E_{iq}\}$ of discs converging h-0-regularly to E_{0q} such that (a) for each i, (1) E_{iq} is in M_i and $E_{iq} \cap K_i = \text{bdry}\, E_{iq}$ for $q \neq 0, m$, (2) $E_{i0} \cup E_{im} \subset K_i$, (3) E_{iq} separates E_{iq-1} from E_{iq+1} in M_i for $q \neq 0, m$, and (4) $E_{iq} \cap D_{ip}$ is an arc for each p and (b) each sequence $\{E_{iq} \cap D_{ip}\}_i$ converges 0-regularly to $E_{0q} \cap D_{0p}$. Denote by R_{ipq} the closure of the component of $M_i - \cup (D_{ip} \cup E_{ip})$ whose boundary contains discs in E_{iq-1}, E_{iq}, D_{ip-1} and D_{ip}. The sequence $\{R_{ipq}\}_i$ converges h-2-regularly to R_{opq}.

If Theorem 3.10 is now applied to each sequence $\{R_{ipq}\}_i$, it is shown that for each r, there is a sequence $\{F_{ir}\}$ of discs converging h-0-regularly to F_{0r} such that (a) for each i, (1) F_{ir} is in M_i and $F_{ir} \cap K_i = \text{bdry}\, F_{ir}$ for $r \neq 0, m$, (2) $F_{i0} \cup F_{im} \subset K_i$, (3) F_{ir} separates F_{ir-1} from F_{ir+1} in M_i for $r \neq 0, m$

and (4) for each p, q, $E_{iq} \cap F_{ir}$ and $D_{ip} \cap E_{iq} \cap F_{ir}$ is a point and (b) each sequence $\{E_{iq} \cap F_{ir}\}$ and $\{D_{ip} \cap F_{ir}\}_i$ converges h-0-regularly to $E_{0q} \cap F_{0r}$ and $D_{0p} \cap F_{0r}$. Denote by R_{ipqr} the closure of that component of

$$M_i - \cup (D_{ip} \cup E_{ip} \cup F_{ip})$$

whose boundary contains discs in D_{ip-1}, D_{ip}, E_{iq-1}, E_{iq}, F_{ir-1}, and F_{ir}. The sequence $\{R_{ipqr}\}_i$ converges h-2-regularly to R_{0pqr}.

Since every sequence of discs which converges 0-regularly to a disc converges completely regularly [6], it follows that for sufficiently large i, there is an $\epsilon/2$-homeomorphism h_i of $\cup (D_{0p} \cup E_{0p} \cup F_{0p})$ onto $\cup (D_{ip} \cup E_{ip} \cup F_{ip})$. For sufficiently large i, it is also true that R_{ipqr} lies in an $\epsilon/2$-neighborhood of R_{0pqr} and has diameter less than $\epsilon/2$. A homeomorphism g_i of M_0 onto M_i which, for each p, q, r, extends $h_i \mid \operatorname{bdry} R_{0pqr}$ to a homeomorphism of R_{0pqr} onto R_{ipqr} is an ϵ-homeomorphism. Thus Theorem 3.12 is proved.

A direct consequence of Theorem 3.12 is the main theorem of this section.

THEOREM 3.13. *If f is a homotopy 2-regular mapping of a metric space X onto a metric space Y such that each inverse under f is a 3-cell, then f is completely regular.*

4. Regular mappings whose inverses are 3-manifolds. In this section, the notation of *section* 2 is used. Each M_i is a compact 3-manifold with boundary imbeddable in E^3 and the sequence $\{K_i\}$ converges completely regularly to K_0. Let $\{\epsilon_i\}$ be a sequence of positive numbers converging to 0 and $\{g_i\}$ be a sequence of mappings such that for each i, g_i is a piecewise linear ϵ_i homeomorphism of K_0 onto K_i.

LEMMA 4.1. *If D_0 is a disc in M_0 with boundary J_0 such that $D_0 \cap K_0 = J_0$, then there is a sequence $\{D_i\}$ of discs converging completely regularly to D_0 whose boundaries $\{J_i\}$ are such that for each i, $D_i \subset M_i$ and $D_i \cap K_i = J_i = g_i(J_0)$.*

Proof. If ϵ is a positive number, there is a regular ϵ-neighborhood A_0 of J_0 in K_0 which is an annulus bounded by simple closed curves J_0' and J_0''. Also, there are discs D_0' and D_0'' in M_0 such that $D_0' \cap K_0 = \operatorname{bdry} D_0' = J_0'$, $D_0'' \cap K_0 = \operatorname{bdry} D_0'' = J_0''$, and the 2-sphere $D_0' \cup A_0 \cup D_0''$ bounds a 3-cell C_0 in M_0 which is a regular ϵ-neighborhood of D_0. It follows from Lemma 2.13 that there are sequences of 2-cells, $\{D_i'\}$ and $\{D_i''\}$ converging strongly to D_0' and D_0'' such that for each i, $D_i' \cup D_i'' \subset M_i$, $D_i' \cap K_i = \operatorname{bdry} D_i' = g_i(J_0')$ and $D_0'' \cap K_i = \operatorname{bdry} D_0'' = g_i(J_0'')$. It follows from Lemma 2.15

that for sufficiently large i, the 2-sphere $D_i' \cup g_i(A_0) \cup D_i''$ bounds a 3-cell C_i in M_i and the sequences $\{C_i\}$ and $\{M_i - C_i\}$ converge to C_0 and $\mathrm{cl}(M_0 - C_0)$. Since the sequence $\{g_i(A_0)\}$ converges completely regularly to A_0 and the convergence of $\{C_i\}$ is h-2-regular at each point of $\mathrm{int}\, C_0$, the lemma and theorems of Section 3 up to and including Theorem 3.10 can now be applied to arcs, discs and annuli in C_0 whose boundaries lie in A_0. This proves Lemma 4.1.

LEMMA 4.2. *If A_0 is an annulus in M_0 with boundary curves J_0' and J_0'' such that $A_0 \cap K_0 = J_0' \cup J_0''$, then there is a sequence $\{A_i\}$ of annuli converging completely regularly to A_0 such that for each i, $K_i \cap A_i = \mathrm{bdry}\, A_i = g_i(J_0' \cup J_0'')$.*

Proof. If ϵ is a positive number, there are regular ϵ-neighborhoods A_0' and A_0'' of J_0' and J_0'' in K_0 which are annuli with boundary curves Z_0', X_0' and Z_0'', X_0'' respectively. Also, there are annuli B_0 and C_0 in M_0 such that $B_0 \cap K_0 = \mathrm{bdry}\, B_0 = X_0' \cup X_0''$, $C_0 \cap K_0 = \mathrm{bdry}\, C_0 = Z_0' \cup Z_0''$, and the torus $B_0 \cup C_0 \cup A_0' \cup A_0''$ bounds a 3-manifold with boundary, T_0, in M_0 which contains A_0 in its interior. It follows from a slight extension of Lemma 2.13 that there are sequences of annuli $\{B_i\}$ and $\{C_i\}$ converging strongly to B_0 and C_0 such that for each i, $B_i \cup C_i \subset M_i$, $B_i \cap K_i = \mathrm{bdry}\, B_i = g_i(X_0 \cup X_0'')$ and $C_i \cap K_i = \mathrm{bdry}\, C_i = g_i(Z_0' \cup Z_0'')$. As in the proof of Lemma 2.15, for sufficiently large i, the torus $C_i \cup B_i \cup g_i(A_0' \cup A_0'')$ bounds a compact 3-manifold with boundary, T_i in M_i and the sequences $\{T_i\}$ and $\{M_i - T_i\}$ converge to T_0 and $\mathrm{cl}(M_0 - T_0)$.

If t is an arc in T_i such that $t \cap \mathrm{bdry}\, T_i$ is the union of the endpoints of t and one of these lies in $g_i(A_0')$, the other in $g_i(A_0'')$, then t will be said to be unknotted in T_i provided that it lies in an annulus A_i^* in T_i such that $A_i^* \cap \mathrm{bdry}\, T_i = \mathrm{bdry}\, A_i^*$, $g_i(A_0') \cap A_i^*$ is deformable into $g_i(X_0')$ (isotopically) in $g_0(A_0')$ and $A_i^* \cap g_i(A_0'')$ is isotopically deformable into $g_i(X_0'')$ in $g_i(A_0'')$. With this definition and the fact that $\{g_i(A_0' \cup A_0'')\}$ converges completely regularly to $A_0' \cup A^{*\prime\prime}$, the proofs in Section 3 up to and including that of Theorem 3.10 can be adapted to yield a proof of Lemma 4.2.

LEMMA 4.3. *If x_0 is a 3-cell in M_0 such that $x_0 \cap K_0$ is a compact 2-manifold with boundary ($M_0 - \mathrm{int}\, x_0$ is thus a compact 3-manifold with boundary), then there is a sequence $x_1, x_2, \cdots$ of 3-cells converging completely regularly to x_0 such that for each i, x_i lies in M_i and $x_i \cap K_i = g_i(x_0 \cap K_0)$ and $\{M_i - \mathrm{int}\, x_i\}$ converges h-2-regularly to $M_0 - \mathrm{int}\, x_0$ and*

$$\{\mathrm{bdry}(M_i - \mathrm{int}\, x_i)\}$$

converges completely regularly to $\mathrm{bdry}(M_0 - \mathrm{int}\, x_0)$.

Proof. The lemma is proved by induction on the number of components of $x_0 \cap K_0$. Suppose, first, that $x_0 \cap K_0$ is connected. Then the closure, E, of each component of $\mathrm{bdry}\, x_0 - (x_0 \cap K_0)$ is a disc bounded by a simple closed curve J. Denote these by $E_1, J_1, \cdots, E_k, J_k$. Then it follows from Lemma 4.1 that for each j there exists a sequence $\{E_{ij}\}_i$ of discs converging completely regularly to E_i such that for each i, $E_{ij} \subset M_i$ and $E_{ij} \cap K_i = \mathrm{bdry}\, E_{ij} = g_i(J_j)$. For sufficiently large i, the 2-sphere, $g_i(x_0 \cap K_0) \cup \cup E_{ij}$ bounds a 3-cell x_i in M_i. The sequences $\{x_i\}$ and $\{M_i - \mathrm{int}\, x_i\}$ converge to x_0 and $M_0 - \mathrm{int}\, x_0$. That the convergence is h-2-regular, and hence completely regular, for $\{x_i\}$ now follows from the complete regularity of the convergence of $\{\mathrm{bdry}\, x_i\}$ to $\mathrm{bdry}\, x_0$ and an application of Lemma 3.11, which applies to this case as well as the case in which each M_i is a 3-cell.

Assume Lemma 4.3 to be true for all x_0, M_0 and K_0 for which $x_0 \cap K_0$ has fewer than r components and suppose that $x_0 \cap K_0$ here has r components. There are mutually exclusive simple closed curves J_0' and J_0'' in $\mathrm{bdry}\, x_0 - (x_0 \cap K_0)$ bounding mutually exclusive discs D_0' and D_0'' whose interiors lie in $\mathrm{int}\, x_0$ such that each of D_0' and D_0'' separates $x_0 \cap K_0$ in x_0 and $J_0' \cup J_0''$ bounds an annulus A_0 in $\mathrm{bdry}\, x_0 - (x_0 \cap K_0)$. Denote by x_0', x_0'' and y_0 the 3-cells which are the closures of the sets into which $D_0' \cup D_0''$ separates x_0, D_0' and D_0'' belonging to x_0' and x_0'' respectively, $D_0' \cup D_0''$ belonging to y_0. Each of $x_0' \cap K_0$ and $x_0'' \cap K_0$ has fewer than r components so that there are sequences $\{x_i'\}$ and $\{x_i''\}$ of 3-cells converging completely regularly to x_0' and x_0'' such that for each i, $x_i' \cup x_i'' \subset M_i$ and $(x_i' \cup x_i'') \cap K_i = g_i[(x_0' \cup x_0'') \cap K_0]$ and the sequence $\{M_i - \mathrm{int}(x_i' \cup x_i'')\}$ converges h-2-regularly to $M_0 - \mathrm{int}(x_0' \cup x_0'')$, the convergence of these sets being completely regular on their boundaries. This can be done by first constructing the x_i' and then applying the lemma to $M_0 - \mathrm{int}\, x_0'$ and x_0''.

Extend g_i to a δ_i-homeomorphism g_i' of $K_0 \cup x_0' \cup x_0''$ onto $K_i \cup x_i' \cup x_i''$, the sequence $\{\delta_i\}$ converging to 0. It follows from Lemma 4.2 applied to $M_0 - \mathrm{int}(x_0' \cup x_0'')$ that there is a sequence of annuli, $\{A_i\}$ converging compeletely regularly to A_0 such that for each i, $\mathrm{int}\, A_i \subset M_i - (x_i' \cup x_i'')$ and $\mathrm{bdry}\, A_i = g_i'(J_0' \cup J_0'')$. For sufficiently large i,

$$(\mathrm{bdry}\, x_i' \cup \mathrm{bdry}\, x_i'' \cup A_i) - \mathrm{int}\, g_i'(D_0' \cup D_0'')$$

bounds a 3-cell x_i in M_i and the sequence $\{x_i\}$ satisfies the conditions required of it by the lemma.

THEOREM 4.4. *The sequence $\{M_i\}$ converges to M_0 completely regularly.*

Proof. There is a cellular decomposition G of M_0 in the sense that each

element of G is a polyhedral 3-cell and the intersection of each element of G with the union of any number of elements of the collection consisting of K_0 and the elements of G is a compact 2-manifold with boundary, which may be empty. Such a decomposition is described by Bing in [2], p. 17. If x_0 is an element of G, then $M_0 - \operatorname{int} x_0$ is a compact 3-manifold with boundary, perhaps not connected, and the decomposition G^* of $M_0 - \operatorname{int} x_0$ consisting of the elements of $G - x_0$ is a cellular decomposition in the sense described above.

Arrange the elements of G in a sequence $x_1, x_2, \cdots, x_k$ with the property that for each j, x_j intersects the boundary of $M_0 - \cup x_r$, $r < j$. Since G is a cellular decomposition, this is possible. There is a sequence $\{x_{1i}\}$ of 3-cells converging completely regularly to x_1 such that for each i, $x_{1i} \cap K_i = g_i(x_1 \cap K_0)$ and $x_{1i} \subset M_i$. Extend g_i to an ϵ_{1i}-homeomorphism g_{1i} of $K_0 \cup x_1$ onto $K_i \cup x_{1i}$, the sequence $\{\epsilon_{1i}\}$ converging to 0. The sequence $\{M_i - \operatorname{int} x_{1i}\}$ converges h-2-regularly to $M_0 - \operatorname{int} x_0$ and since $x_2, \cdots, x_k$ is a cellular decomposition of $M_0 - \operatorname{int} x_0$, the above process may be applied to $M_i - \operatorname{int} x_{1i}$ to obtain a sequence $\{x_{2i}\}$ of 3-cells converging completely regular to x_2 such that for each i, $x_{2i} \subset M_i - \operatorname{int} x_{1i}$ and

$$x_{2i} \cap \operatorname{bdry}(M_i - \operatorname{int} x_{ij}) = g_{1i}(x_2 \cap \operatorname{bdry}(M_0 - \operatorname{int} x_1)).$$

The homeomorphism g_{1i} may be extended to an ϵ_{2i}-homeomorphism g_{2i} of $K_0 \cup x_1 \cup x_2$ onto $K_i \cup x_{1i} \cup x_{2i}$. The theorem is now proved by repeating this process for each $j \leqq k$.

A direct consequence of this theorem is the main theorem of this section.

THEOREM 4.5. *If f is an h-2-regular mapping of a metric space X onto a metric space Y such that each inverse under f is a compact 3-manifold with boundary which is imbeddable in E_3 and the boundaries of the inverses under f are mutually homeomorphic, then f is completely regular.*

5. Some consequence of the preceding theory. In [6], Theorem 2, there was proved in a silghtly more general form the

THEOREM A. *Suppose that X is a complete metric space, Y is a metric space with finite covering dimension and f is a completely regular mapping of X onto Y such that (1) for each point p of Y there is a homeomorphism f_p of the $(i+1)$-cell R^{i+1}, with boundary S^i, onto $f^{-1}(p)$ and (2) there is a homeomorphism h of $\cup f_p(S^i)$, $p \in Y$, onto the direct product $Y \times S^i$ such that the diagram*

$$\begin{tikzcd} \cup f_p(S^i) \arrow[r, "h"] \arrow[dr, "f"'] & Y \times S^i \arrow[d, "\pi"] \\ & Y, \end{tikzcd}$$

where π is the projection map, is commutative. Then there is a homeomorphism h^* of X onto the direct product $Y \times R^{i+1}$ which extends h and is such that the diagram

$$\begin{tikzcd} X \arrow[r, "h^*"] \arrow[dr, "f"'] & Y \times R^{i+1} \arrow[d, "\pi"] \\ & Y \end{tikzcd}$$

is commutative.

This theorem and its proof yield the following typical result from [7].

THEOREM B. *If f is a 0-regular mapping of the complete metric space X onto the finite (covering) dimensional space Y such that each inverse f is homeomorphic to the compact 2-manifold with boundary, M, then (X, f, y) is a locally trivial fibre space. If Y is locally compact, separable and contractible, then X is homeomorphic to $Y \times M$, where f corresponds to the projection mappings of $Y \times M$ into Y.*

A consequence of Theorem A and Theorems 3.13 and 4.5 is

THEOREM 5.1. *If f is an h-2-regular mapping of the complete metric space X onto the finite (covering) dimensional space Y such that each inverse under f is a 3-cell, R^3, with boundary S^2, then (X, f, Y) is a locally trivial fibre space. If Y is locally compact, separable and contractible, then X is homeomorphic to $Y \times R^3$, where f corresponds to the projection map of $Y \times R^3$ onto Y. If there is a homeomorphism h of $\cup$ bdry $f^{-1}(p)$, $p \in Y$, onto $Y \times S^2$ such that the diagram*

$$\begin{tikzcd} \cup \operatorname{bdry} f^{-1}(p) \arrow[r, "h"] \arrow[dr, "f"'] & Y \times S^2 \arrow[d, "\pi"] \\ & Y \end{tikzcd}$$

is commutative, then h may be extended to a homeomorphism h^* of X onto Y such that the diagram

$$\begin{tikzcd} X \arrow[r, "h^*"] \arrow[dr, "f"'] & Y \times R^3 \arrow[d, "\pi"] \\ & Y \end{tikzcd}$$

is commutative.

Proof. The mapping f is completely regular. Thus $f \mid \cup \operatorname{bdry} f^{-1}(p)$ is completely regular and $(\cup \operatorname{bdry} f^{-1}(p), f, Y)$ is, by Theorem B, a locally trivial fibre space. The first part of the theorem now follows from Theorem A, as does the third part. If Y is locally compact, separable and contractible, then Theorem B implies that the hypothesis for the third part of the present theorem is fulfilled. Hence X is homeomorphic to $Y \times R^3$, f corresponding to the projection map.

The proof of Theorem A depends strongly on the fact that the space of homeomorphisms of a 3-cell onto itself leaving its boundary pointwise fixed is LC^n for each n[1]. (A space X is LC^n if for each point x in X and each $\epsilon > 0$ there is a $\delta > 0$ such that every mapping of a k-sphere, $k \leqq n$, into $S(x, \delta)$ is homotopic to 0 in $S(x, \epsilon)$.) It is clear from the proof of Theorem A that if M is a compact 3-manifold with boundary and (1) the space of homeomorphisms of M onto itself leaving bdry M pointwise fixed is locally connected (LC^0) and (2) for each positive number ϵ there is a positive number δ such that every δ-homeomorphism of bdry M onto itself can be extended to an ϵ-homeomorphism of M onto itself, then Theorem A remains true for one-dimensional Y if R^{i+1} is replaced by M and S^i is replaced by the boundary of M. (See the proof of Theorem 3 of [7].) Proofs of these two facts are included here. The proofs of Lemmas 5.4 and 5.3 were suggested by J. H. Roberts [13]. Lemma 5.4 has been proved by Sanderson. His proof is not yet published, but see [14]. For further results, see [8]. See also the recent work of Kister and Fisher to appear in the Transactions of the Ameriacn Mathematical Society ([3], [4], and [9]).

LEMMA 5.2. *If M is a compact 3-manifold with boundary, then for each positive number ϵ there is a positive number δ such that every δ-homeomorphism of* bdry M *onto itself can be extended to an ϵ-homeomorphism of M onto itself.*

Proof. Denote by $C_1, C_2, \cdots$ the components of the boundary of M. Each C_i is a compact 2-manifold. There is a homeomorphism h of bdry $M \times I$, where I is the unit interval, into M such that for each point p in bdry M,

$h(p, I)$ has diameter less than $\epsilon/3$ and $h(p,1) = p$. (See [11].) It follows from Theorem 1 of [7] that the space of homeomorphisms of bdry M onto itself is locally connected. Denote by H this space of homeomorphisms and by i its identity. There is a positive number δ such that if $f \in S(i,\delta)$ in H, then there is a mapping F of I into $S(i,\epsilon/3)$ such that $F(0) = i$ and $F(1) = f$. If $y \in I$ and $q = h(p,y)$, let $f^*(q)$ denote $h(F(y)(p),y)$. If $q \in M - h(\text{bdry}\, M \times I)$, let $f^*(q) = q$. Since $d(q,p) < \epsilon/3$, $d(p, F(y)(p)) < \epsilon/3$ and $d[F(y)(p), h(F(y)(p),y)] < \epsilon/3$, and $f^*(q) = h(F(1)(q),1) = h(f(q),1) = f(q)$. Thus f^* extends f and the lemma is proved.

LEMMA 5.3. *Suppose that K is a 3-manifold with boundary, K_2 is a polyhedral 3-cell in K which intersects* bdry K *in a 2-manifold or not at all and K_2' is a polyhedral 3-cell in K_2 such that $K_2' \cap \text{bdry}\, K_2 \subset \text{int}(K_2 \cap \text{bdry}\, K)$. If ϵ is a positive number, then there is a positive number δ such that if f is a piecewise linear δ-homemorphism of K onto itself leaving* bdry K *pointwise fixed, then there is a piecewise linear ϵ-homeomorphism g^* of K onto itself which is the identity outside K_2, is g on K_2' and leaves* bdry K *pointwise fixed.*

Proof. Denote by K^* the compact 3-manifold obtained by "sewing" a homeomorphic image of K to K along its boundary. Let K_1 denote $\text{cl}(K - K_2)$ and U a polyhedral neighborhood of $K_1 \cap K_2$ in K such that $U \cap K_2' = 0$, U is a 3-manifold with boundary, $\text{cl}(\text{int}\, K \cap \text{bdry}\, U)$ is a 2-manifold with boundary and $U \cap K_2$ is a 3-manifold with boundary. Let V denote a polyhedral neighborhood of $\text{bdry}(U \cap K_2)$ in K^* such that $V \cap K_2' = 0$. If ϵ is a positive number, it follows from Moise's lemma on the fitting together of homeomorphisms [11] that there is a positive number δ such that if t_1 is a piecewise linear δ-homomorphism of $(U \cap K_2) \cup V$ into K^* and t_2 is a piecewise linear δ-homeomorphism of $K^* - (U \cap K_2)$ into K^*, then there is a piecewise linear ϵ-homeomorphism t of K^* onto itself such that $t \mid (K_2 \cap \text{U} - V = t_1$ and $t \mid K^* - (U \cap K_2) = t_2$.

Suppose that g is a piecewise linear δ-homeomorphism of K onto itself (hence into K^*) leaving bdry K pointwise fixed. Denote by t_1 the identity homeomorphism of $(U \cap K_2) \cup V$ into K^* and by t_2 the δ-homeomorphism of $K^* - (U \cap K_2)$ into K^* which is identical to g when restricted to $K_2 - (U \cap K_2)$ and is the identity when restricted to $K_1 \cup (K^* - K)$. Then there is a piecewise linear ϵ-homeomorphism f of K^* onto itself such that $f \mid (U \cap K_2) - V$ is the identity and $f \mid K^* - (U \cap K_2) = t_2$. Thus $g^* = f \mid K$ is such that $g^* \mid K_2' = g \mid K_2'$ and $g^* \mid K_1$ is the identity, since bdry $K \subset \text{cl}(K^* - (K_2 \cap U)$, and $g^* \mid \text{bdry}\, K$ is the identity on $K_1 \cup (K^* - K)$.

LEMMA 5.4. *If K is a compact 3-manifold with boundary, then the space of homeomorphisms of K onto itself leaving* bdry K *pointwise fixed is locally connected.*

Proof. Let G be a cellular decomposition of K as described in Section 4. Induction will be used on the number of elements of G. If G has just one element, then K is a 3-cell and the lemma follows from a well known theorem of Alexander [1]. Suppose the lemma to be true for all compact 3-manifolds with boundary which have a cellular decomposition with fewer than k elements and that G, here, has k elements. Denote by X an element of G and by X' a 3-cell in X such that $X' \cap \operatorname{bdry} X \subset \operatorname{int}(X \cap \operatorname{bdry} K)$ and $\operatorname{cl}(K - X')$ is homeomorphic to $\operatorname{cl}(K - X)$. Denote by $H(K)$, $H(X)$, and $H(\operatorname{cl}(K - X'))$ the spaces of homeomorphisms of K, X, and $\operatorname{cl}(K - X')$ onto themselves leaving the boundaries pointwise fixed. (If f and g are homeomorphisms in one of these spaces, $d(f, g) = \operatorname{lub}[d(f(x), g(x))]$.) Suppose ϵ is a positive number. There is a positive number δ' such that (1) if f is a $2\delta'$-homeomorphism in $H(X)$, there is a mapping F of I into $H(X)$ such that $F(0) = \mathbf{i} \mid X$, $F(1) = f$ and $F(t)$ move no point as much as $\epsilon/2$ and (2) if f is a $2\delta'$-mapping in $H(\operatorname{cl}(K - X'))$, there is a mapping F of I into $H(\operatorname{cl}(K - X'))$ such that $F(0) = \mathbf{i} \mid \operatorname{cl}(K - X')$, $F(1) = f$ and $F(t)$ moves no point as much as $\epsilon/2$. Statement (2) follows from the induction hypothesis and the fact that the elements of $G - (X)$ form a cellular decomposition of $\operatorname{cl}(K - X)$, which is homeomorphic to $\operatorname{cl}(K - X')$, with fewer than k elements. Also, there exists a positive number δ such that if g is a piecewise linear δ-homeomorphism in $H(K)$, then there is a piecewise linear δ'-homeomirphism g^* such that $g^* \mid X' = g \mid X'$ and $g^* \mid K - X = \mathbf{i} \mid K - X$.

Suppose then that g is a piecewise linear δ-homeomorphism in $H(K)$. There is a δ'-homeomorphism g^* such that $g^* \mid X' = g \mid X'$ and $g^* \mid K - X = \mathbf{i} \mid K - X$. Clearly $g^* \mid X$ is an element of $H(X)$. Thus there is a mapping F of I into $H(K)$ such that $F(0) = \mathbf{i}$, $F(1) = g^*$ and $F(t)$ is an $\epsilon/2$-homeomorphism which moves no point of $K - X$. The mapping $g^{*-1}g \mid X' = \mathbf{i} \mid X'$ and therefore $g^{*-1}g \mid \operatorname{cl}(K - X')$ is an element of $H(\operatorname{cl}(K - X'))$ and moves no point as much as $2\delta'$. Hence there is a mapping F^* of I into $H(K)$ such that $F^*(0) = i$, $F^*(1) = g^{*-1}g$ and $F^*(t)$ is an $\epsilon/2$-homeomorphism which moves no point of X'. Let $Z(t)$ denote $F(t)F^*(t)$. Then $Z(0) = \mathbf{i}$ and $Z(1) = g^*g^{*-1}g = g$. Furthermore, each $Z(t)$ is an ϵ-homeomorphism. Thus each piecewise linear δ-homeomorphism is connected to the identity by an arc of diameter less than 2ϵ.

Now let g be any $\delta/2$-homeomorphism in $H(K)$ and let $\{\delta_i\}$ and $\{\epsilon_i\}$ be

decreasing sequences of positive numbers converging to 0 such that each piecewise linear $2\delta_i$-homeomorphism in $H(K)$ may be joined to the identity by an arc of ϵ_i-homeomorphisms. Let K^* be as described in the proof of Lemma 5.3 and let U denote a polyhedral neighborhood of $\operatorname{bdry} K$ in K^*. Denote by g^* the homeomorphism of K^* onto itself such that $g^* \mid K = g$ and $g^* \mid K^* - K = i$. From the lemma on the fitting together of homeomorphisms [11] it follows that there is a positive number δ_i' such that if f_i and f_i' are piecewise linear δ_i'-approximations to $g^* \mid K \cup U$ and $g^* \mid \operatorname{cl}(K^* - K)$, then there is a piecewise linear δ_i-approximation, g_i^* to g^* such that $g^* \mid K^* - K = f_i'$ and $g_i^* \mid K - U = f_i$. From the theorem of Moise on the approximation of homeomorphisms by piecewise linear ones [11] it follows that there is a piecewise linear δ_i-approximation (homeomorphism) f_i to $g^* \mid K \cup U$. Denote by f_i' the mapping $i \mid \operatorname{cl}(K^* - K)$. Then there is a piecewise linear δ_i-approximation g_i^* to g^* such that $g_i^* \mid K^* - K = i$. Then $g_i = g_i^* \mid K$ is an element of $H(K)$ and is a piecewise linear δ_i-approximation to g.

It may be assumed that $d(g_i, g_{i+1}) < 2\delta_i$. Then $d(i, g_{i+1}g_i^{-1}) < 2\delta_i$, so that there is a mapping F_i of I into $H(K)$ such that $F_i(0) = i$ and $F_i(1) = g_{i+1}g_i^{-1}$ and $d(i, F_i(t)) < \epsilon_i$. Then the mapping F_i^* of I into $H(K)$ defined by the equation $F_i^*(t) = F_i(t)g_i$ is such that $F_{i0}(0) = g_i$, $F_i^*(1) = g_{i+1}$ and $d(g_i, F_i^*(t)) < \epsilon_i$. Since the sequence $\{g_i\}$ converges to g, $\{\epsilon_i\}$ converges to 0 and there is an arc of diameter less than $2\epsilon_i$ connecting g_i to g_{i+1}, it has been shown that there is an arc of diameter less than $\delta/2$ connecting g to a piecewise linear homeomorphism g' in $H(K)$. Since g' can be connected to i by an arc of diameter less than 2ϵ, the local connectedness of $H(K)$ at the identity is established. Since $H(K)$ can be given a group structure, it is locally connected at all points and the lemma is proved.

THEOREM 5.5. *If f is an h-2-regular mapping of a complete metric space X onto a one-dimensional space Y such that each inverse under f is homeomorphic to the compact 3-manifold M with boundary which is imbeddable in E^3, then (X, f, Y) is a locally trivial fibre space. If Y is locally compact, separable and contractible, then X is homeomorphic to $M \times Y$, where f corresponds to the projection map of $M \times Y$ onto Y. If there is a homeomorphism f of* $\cup \operatorname{bdry} f^{-1}(p)$, $p \in Y$, *onto* $\operatorname{bdry} M \times Y$ *such that the diagram*

$$
\begin{tikzcd}
\cup \operatorname{bdry} f^{-1}(p) \arrow[r, "h"] \arrow[dr, "f"'] & \operatorname{bdry} M \times Y \arrow[d, "\pi"] \\
& Y
\end{tikzcd}
$$

is commutative, then, if the space of homeomorphisms of M onto itself leaving its boundary pointwise fixed is connected, h may be extended to a homeomorphism h^* of X onto $M \times Y$ such that the diagram

$$\begin{tikzcd} X \arrow[r, "h^*"] \arrow[dr, "f"'] & M \times Y \arrow[d, "\pi"] \\ & Y \end{tikzcd}$$

is commutative.

Proof. It follows from Theorem 4.5 that f is completely regular. Thus, by Theorem B, $(\cup \operatorname{bdry} f^{-1}(p), f, Y)$ is a locally trivial fibre space. The theorem now follows from Lemmas 5.2 and 5.4 and a slight modification of the proof of Theorem A (see Thesorem 3 of [7]).

Note. In view of Theorem 4.5, if Y is connected, it is only necessary to assume here that each $f^{-1}(p)$ is homeomorphic to a 3-manifold with boundary which is imbeddable in E^3 and that the boundaries of the inverses are mutually homeomorphic. Also, the restriction on Y that it be one-dimensional is now known to be unnecessary. If the space of homeomorphisms of M onto itself leaving $\operatorname{bdry} M$ pointwise fixed is LC^n, then Y may be taken to be $(n+1)$-dimensional, as the proof of Theorem A indicates. That this is true for each n will be proved in a later paper [8].

6. Some weakening of the hypotheses in earlier theorems. In this section, the notation of section 2, unless specific modification are stated, will be used.

THEOREM 6.1. *If the sequence $\{M_i\}$ converges to M_0 h-1-regularly, then it converges h-2-regularly.*

Proof. Note first that none of the proofs in Section 3 require more than h-1-regularity. Suppose that P is a point in $\operatorname{int} M_0$ and that ϵ is a positive number. Let S_0 denote a 2-sphere in $\operatorname{int} M_0$ bounding a 3-cell in $\operatorname{int} M_0$ whose interior contains P and has diameter less than ϵ. Then it follows from Lemma 2.15 that there is a sequence $\{S_i\}$ of 2-spheres converging strongly to S_0 such that for each i, S_i bounds a 3-cell A_i in M_i and the sequences $\{A_i\}$ and $\{M_i - \operatorname{int} A_i\}$ converges to A_0 and $M_0 - \operatorname{int} A_0$. For sufficiently large i, A_i is a subset of $S(P, \epsilon)$. Let δ be a positive number such that for each i,

$S(P,\delta) \cap M_i \subset \operatorname{int} A_i$. Then every mapping of a 2-sphere into $S(P,\delta) \cap M_i$ is homotopic to 0 in A_i and consequently in $S(P,\epsilon) \cap M_i$.

If P is a point of K_0, let C_0 denote a disc in K_0 whose interior contains P and has diameter less than ϵ. Let D_0 denote a disc in M_0 such that $D_0 \cap K_0 = \operatorname{bdry} D_0 = \operatorname{bdry} C_0$ and the 2-sphere $D_0 \cup C_0$ bounds a 3-cell in M_0 of diameter less than ϵ. From Lemma 2.13 and Theorem 2.14 it follows that there is a sequence $\{D_i\}$ of discs converging to D_0 such that for each i, $\{D_i\}$ lies in M_i, $D_i \cap K_i = \operatorname{bdry} D_i$ which is the boundary of a disc C_i in K_i, $C_i \cup D_i$ bounds a 3-cell A_i in M_i and the sequences $\{A_i\}$ and $\{M_i - \operatorname{int} A_i\}$ converge to A_0 and $M_0 - \operatorname{int} A_0$. A repetition of the argument in the foregoing paragraph now demonstrates that the convergence is h-2-regular at each point of K_0 and Theorem 6.1 is proved.

THEOREM 6.2. *If the sequence $\{M_i\}$ of compact 3-manifolds with boundary converges h-2-regularly to the compact 3-manifold with boundary M_0 and each M_i is imbeddable in E^3, then for sufficiently large i,* $\operatorname{bdry} M_i$ *is homeomorphic to* $\operatorname{bdry} M_0$.

Proof. As before, denote $\operatorname{bdry} M_i$ by K_i. It follows from Lemma 2.9 that if C_0 is a component of K_0, then there is a sequence $\{C_i\}$ of components of $\{K_i\}$ converging strongly to C_0. It will first be proved that this convergence is completely regular. Suppose that J_0 is a simple closed curve bounding a disc A_0 in C_0. The argument for Theorem 2.14 may be applied to prove that there is a sequence $\{J_i\}$ of simple closed curves converging strongly to J_0 such that for each i, J_i bounds a compact 2-manifold with boundary, A_i, in C_i, the sequences $\{A_i\}$ and $\{C_i - \operatorname{int} A_i\}$ converging to A_0 and $C_0 - \operatorname{int} A_0$. Suppose that A_i is a disc with handles. Then there is a pair of simple closed curves, x_i and y_i, in A_i which cross each other and have only one point in common. It follows from the 1-regularity of the convergence of M_i to M_0 that if A_0 has sufficiently small diameter, x_i and y_i bound singular discs B_0 and D_i in M_i. Dehn's Lemma implies that B_i and D_i may be taken to be non-singular and such that $B_i \cap K_i = x_i$ and $D_i \cap K_i = y_i$. Small adjustments may be made in D_i and B_i so that each component of $B_i \cap D_i$ other than $x_i \cap y_i$ is a simple closed curve. These components may be removed in a manner described earlier in this paper. This process leaves two discs B_i' and D_i' such that $B_i' \cap D_i' = x_i \cap y_i$, $B_i' \cap K_i = x_i$ and $D_i' \cap K_i = y_i$. This is clearly impossible. Hence for sufficiently large i, A_i is a disc. It follows from the proof of Theorem 2.14 that the convergence of $\{C_i\}$ to C_0 is h-1-regular and hence completely regular.

Theorem 6.2 will now be proved when it is shown that no point of C_0

is a limit point of $\cup(K_i - C_i)$ and that no point of $\operatorname{int} M_i$ is a limit point of $\cup K_i$. Let ϵ be a positive number, P a point of C_0 and A_0 a disc in C_0 of diameter less than ϵ whose interior contains P. There is a disc D_0 in M_0 such that $D_0 \cap K_0 = \operatorname{bdry} D_0 = \operatorname{bdry} A_0$ and the 2-sphere $D_0 \cup A_0$ bounds a 3-cell N_0 in M_0 of diameter less than ϵ. It follows from the proofs of Lemma 2.13 and Theorem 2.14 that there are sequences $\{A_i\}$ and $\{D_i\}$ converging to A_0 and D_0 such that for each i, A_i is a disc in C_i, D_i is a disc in M_i, $D_i \cap K_i = \operatorname{bdry} D_i = \operatorname{bdry} A_i$ and that $M_i - (D_i \cup A_i)$ has two components, the closure of one denoted by N_i, such that the sequences $\{N_i\}$ and $\{M_i - \operatorname{int} N_i\}$ converge to subsets of N_0 and $M_0 - \operatorname{int} N_0$. No component of $K_i - C_i$ intersects D_i so that if P is a limit point of $\cup(K_i - C_i)$, then there is a sequence $\{R_{n_i}\}_i$ of components of $\{K_{n_i}\}$ converging to a subset of N_0. However, it follows from the h-2-regularity of the convergence that for sufficiently large i and small ϵ, $D_i \cup A_i$ bounds a singular 3-cell, N_i', in M_i which, since M_i is imbeddable in E^3, contains a non-singular 3-cell, N_i'', in M_i whose boundary is $D_i \cup A_i$. But $R_{n_i} \subset N_{n_i}''$ for large i, which contradicts the fact that $R_i \subset K_i$.

If P is a point of $\operatorname{int} M_0$, let S_0 be a 3-sphere in $\operatorname{int} M_0$ bounding a 3-cell A_0 in $\operatorname{int} M_0$ whose interior contains P and has diameter less than ϵ. It follows from Lemma 2.15 that there is a sequence $\{S_i\}$ of 2-spheres converging strongly to S_0 such that for each i, $S_i \subset \operatorname{int} M_i$ and $M_i - S_i$ has two components, the closure of one denoted by A_i, such that the sequences $\{A_i\}$ and $\{M_i - \operatorname{int} A_i\}$ converge to subsets of A_0 and $M_0 - \operatorname{int} A_0$. The argument in the foregoing paragraph may now be used to prove that P is not a limit point of $\cup K_i$. This completes the proof of Theorem 6.2.

It is easy to find examples to show that Theorem 6.1 is not true if regularity is assumed only in dimensions 0 and 2 and that Theorem 6.2 is not true if only h-1-regularity is assumed. The theorems in this section demonstrate that the mapping f in Theorems 5.1 and 5.5 need only be h-1-regular and that under the hypothesis of h-1-regularity and the connectedness of Y, the first two parts of Theorem 5.5 remain true if each $f^{-1}(p)$ is only assumed to be a compact 3-manifold imbeddable in E^3.

GOUCHER COLLEGE.

REFERENCES.

[1] J. W. Alexander, "On the deformation of an n-cell, *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 9 (1923), pp. 406-407.

[2] R. H. Bing, "A characterization of 3-space by partitionings," *Transactions of the American Mathematical Society*, vol. 70 (1951), pp. 15-27.

[3] G. M. Fisher, "On the group of all homeomorphisms of a manifold," *Notices of the American Mathematical Society*, vol. 6 (1959), p. 543.

[4] ———, "On the group of all homeomorphisms of a manifold," *ibid.*, vol. 6 (1959), p. 808.

[5] R. H. Fox, "On the imbedding of polyhedra in 3-space," *Annals of Mathematics*, vol. 49 (1948), pp. 462-470.

[6] M.-E. Hamstrom and E. Dyer, "Completely regular mappings," *Fundamenta Mathematicae*, vol. 45 (1958), pp. 103-118.

[7] ———, "Regular mappings and the space of homeomorphisms on a 2-manifold," *Duke Mathematical Journal*, vol. 25 (1958), pp. 521-532.

[8] M.-E. Hamstrom, "Regular mappings and the space of homeomorphisms on a 3-manifold," *Notices of the American Mathematical Society*, vol. 6 (1959), p. 783.

[9] J. M. Kister, "Isotopies on 3-manifolds with boundaries, I and II," *ibid.*, vol. 6 (1959), p. 642 and p. 764.

[10] V. L. Klee, Jr., "Some topological properties of convex sets," *Transactions of the American Mathematical Society*, vol. 78 (1955), pp. 30-45.

[11] E. E. Moise, "Affine structures in 3-manifolds, V, The triangulation theorem and hauptvermutung," *Annals of Mathematics*, vol. 56 (1952), pp. 96-114.

[12] C. D. Papakyriakopoulos, "On Dehn's lemma and the asphericity of knots," *ibid.*, vol. 66 (1957), pp. 1-26.

[13] J. H. Roberts, "Local arcwise conectivity in the space H^n of homeomorphisms of S^n onto itself," *Summary of Lectures, Summer Institute on Set Theoretic Topology*, Madison, Wisconsin, 1955, p. 100.

[14] D. E. Sanderson, "On isotopy of homeomorphisms of E^3 and S^3," *Bulletin of the American Mathematical Society*, Abstract 60-4-545.

[15] A. Shapiro and J. H. C. Whitehead, "A proof and extension of Dehn's lemma," *ibid.*, vol. 64 (1958), pp. 174-178.

[16] J. H. C. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society*, vol. 45 (1939), pp. 243-327.

BASIC REPRESENTATIONS OF COMPLETELY SIMPLE SEMIGROUPS.* [1]

By A. H. CLIFFORD.

1. **Introduction.** In a previous paper [1], the author discussed the theory of representations of a completely simple semigroup S by matrices over a field Ω. According to the fundamental theorem of Rees [2], S is isomorphic with, and hence may be taken to be, a regular matrix semigroup over a group with zero. It was shown in [1] that every representation $\mathfrak{T}^*$ of S induces a representation $\mathfrak{T}$ of G; we call $\mathfrak{T}^*$ an extension of $\mathfrak{T}$ to S.

A given representation $\mathfrak{T}$ of G may not be extendible to a representation $\mathfrak{T}^*$ of S; but if it is so extendible, then the extension $\mathfrak{T}_0{}^*$ of $\mathfrak{T}$ of least possible degree over Ω is uniquely determined by $\mathfrak{T}$ to within equivalence. We call $\mathfrak{T}_0{}^*$ the basic extension of $\mathfrak{T}$, and by a basic representation of S we shall mean one that is the basic extension to S of a representation of G. Any extension $\mathfrak{T}^*$ of a representation $\mathfrak{T}$ of G reduces (but does not in general decompose) into the basic extension $\mathfrak{T}_0{}^*$ of $\mathfrak{T}$ and null representations.

It is immediate from Theorems 4.1 and 6.1 of [1] that the mapping $\mathfrak{T} \to \mathfrak{T}_0{}^*$ is one-to-one (in the sense of equivalence) from the extendible representations of G to the basic representations of S. However, several questions concerning this correspondence were left unanswered. It was shown (Theorem 7.1) that if $\mathfrak{T}$ is irreducible, so is $\mathfrak{T}_0{}^*$, but the converse was left open. One of the main purposes of this note is to prove that the converse is true, and hence that *all the irreducible representations of S over Ω are obtained as the basic extensions to S of the extendible irreducible representations of G.*

In § 2 we show that the correspondence $\mathfrak{T} \to \mathfrak{T}_0{}^*$ preserves decomposition. In § 3 we show that it preserves reduction in a limited sense: the *non-null* irreducible constituents of $\mathfrak{T}_0{}^*$ are the basic extensions of the irreducible constituents of $\mathfrak{T}$. An example in § 4 shows that an extraneous null constituent can occur in $\mathfrak{T}_0{}^*$. (Thanks to W. D. Munn for pointing this out.)

* Received July 18, 1959.

[1] This paper was prepared with the partial support of the National Science Foundation grant to the Tulane Mathematics Department.

The author would like to take this opportunity of mentioning that the underlying ideas and methods of [1] should be attributed to Suschkewitsch [3]. He also proved the first part of Theorem 3.1. The remark in the introduction of [1] that Suschkewitsch "made considerable progress" was neither precise nor adequate.

2. Decomposition.

THEOREM 1. *A representation $\mathfrak{X}$ of G is extendible to S if and only if each of its indecomposable constituents is extendible. If $\mathfrak{X}$ is extendible, then the indecomposable constituents of the basic extension $\mathfrak{X}_0{}^*$ of $\mathfrak{X}$ are the basic extensions of the indecomposable constituents of $\mathfrak{X}$. In particular, $\mathfrak{X}_0{}^*$ is indecomposable if and only if $\mathfrak{X}$ is indecomposable.*

Proof. First let $\mathfrak{X}$ be an indecomposable representation of G which is extendible to S, and let $\mathfrak{X}_0{}^*$ be its basic extension to S. Suppose that $\mathfrak{X}_0{}^*$ could be decomposed into two representations $\mathfrak{R}$ and $\mathfrak{R}'$ of S each of lower degree than that of $\mathfrak{X}_0{}^*$. The restrictions of $\mathfrak{R}$ and $\mathfrak{R}'$ to G cannot share the indecomposable representation $\mathfrak{X}$ of G. Hence either $\mathfrak{R}$ or $\mathfrak{R}'$ is an extension to S of $\mathfrak{X}$, contrary to the fact that $\mathfrak{X}_0{}^*$ is the extension of $\mathfrak{X}$ to S of lowest possible degree.

Now suppose that $\mathfrak{X}$ is the direct sum $\mathfrak{X}' \oplus \mathfrak{X}''$ of two representations of G each of lower degree than that of $\mathfrak{X}$. According to Theorem 7.2 of [1], $\mathfrak{X}$ is extendible to S if and only if $\mathfrak{X}'$ and $\mathfrak{X}''$ are both extendible; and, if this is the case, then the basic extension of $\mathfrak{X}$ is equivalent to the direct sum of the basic extensions of $\mathfrak{X}'$ and $\mathfrak{X}''$. The rest of Theorem 1 then follows by an evident induction on the number of indecomposable constituents of $\mathfrak{X}$.

3. Reduction.

THEOREM 2. *Let $\mathfrak{X}$ be an extendible representation of G, and let $\mathfrak{X}^*$ be any extension of $\mathfrak{X}$ to S. Then the non-null irreducible constituents of $\mathfrak{X}^*$ are the basic extensions of the irreducible constituents of $\mathfrak{X}$. The basic extension $\mathfrak{X}_0{}^*$ of $\mathfrak{X}$ is irreducible if and only if $\mathfrak{X}$ is irreducible.*

Proof. Let $\mathfrak{X}$ be an extendible representation of G, and let $\mathfrak{X}^*$ be any extension of $\mathfrak{X}$ to S. Assume that $\mathfrak{X}$ reduces into representations $\mathfrak{X}'$ and $\mathfrak{X}''$ of G, each of lower degree over Ω than that of $\mathfrak{X}$. We proceed to show that $\mathfrak{X}^*$ reduces into two representations of S, one of which is an extension of $\mathfrak{X}'$ and the other of $\mathfrak{X}''$.

Equation (3.1) of [1] shows that the restriction of $\mathfrak{T}^*$ to G decomposes into the proper representation $\mathfrak{T}$ of G and a null representation of G. Let the corersponding decomposition of the representation space V of $\mathfrak{T}$ be $V = V_1 \oplus V_2$, where V_1 carries $\mathfrak{T}$ and V_2 carries the null representation. Let n be the dimension of V_1 over Ω, and t that of V_2. By Theorem 3.1 of [1], we may assume that the representing matrices of $\mathfrak{T}^*$ have the form

$$(3.5) \qquad T^*[(a)_{\iota\kappa}] = \begin{pmatrix} T(p_{1\iota}ap_{\kappa 1}) & T(p_{1\iota}a)Q_\kappa \\ R_\iota T(ap_{\kappa 1}) & R_\iota T(a)Q_\kappa \end{pmatrix},$$

where the R_ι are $t \times n$ matrices and the Q_κ are $n \times t$ matrices satisfying

$$(3.7) \qquad Q_\kappa R_\iota = T(p_{\kappa\iota}) - T(p_{\kappa 1}p_{1\iota}).$$

Let W_1 be the invariant subspace of V_1 which carries $\mathfrak{T}'$, so that $\mathfrak{T}''$ is carried by the factor-space V_1/W_1. Let W be the subspace of V consisting of all vectors w of V having the form

$$w = x + \sum_{\iota \in J} R_\iota x_\iota$$

with x and the x_ι in W_1, and where the sum is finite, that is, all but a finite number of the x_ι are the zero vector of W_1. Now R_ι (for each ι in J) may be regarded as a linear transformation of V_1 into V_2. Thus $w = x + y$ with x in W_1 and $y = \sum_\iota R_\iota x_\iota$ in V_2. It will be convenient in what follows to write w in block form $\begin{pmatrix} x \\ y \end{pmatrix}$ corresponding to (3.5). Since $W \subseteq W_1 \oplus V_2$, it is a proper subspace of V, and we proceed to show that it is invariant under $\mathfrak{T}^*$.

By direct calculation from (3.5), we have

$$T^*[(a)_{\iota\kappa}] \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

where, using (3.7) and $y = \sum_j R_j x_j$,

$$\begin{aligned} x' &= T(p_{1\iota}ap_{\kappa 1})x + T(p_{1\iota}a)\sum_j Q_\kappa R_j x_j \\ &= T(p_{1\iota}ap_{\kappa 1})x + \sum_j [T(p_{1\iota}ap_{\kappa j}) - T(p_{1\iota}ap_{\kappa 1}p_{1j})]x_j \end{aligned}$$

and

$$\begin{aligned} y' &= R_\iota T(ap_{\kappa 1})x + R_\iota T(a)\sum_j Q_\kappa R_j x_j \\ &= R_\iota T(ap_{\kappa 1})x + R_\iota \sum_j [T(ap_{\kappa j}) - T(ap_{\kappa 1}p_{1j})]x_j. \end{aligned}$$

Since x and all the x_ι belong to W_1, and W_1 is invariant under $T(b)$ for

every b in G, it is clear that $x' \in W_1$. Since y' is seen to be of the form $R_i x''$ wtih x'' in W_1, it follows that $x' + y' \in W$, and so W is invariant under $\mathfrak{T}^*$.

Let $\mathfrak{R}'$ be the representation of S carried by the invariant subspace W of V constructed above, and let $\mathfrak{R}''$ be that carried by the factor space V/W, so that $\mathfrak{T}^*$ reduces into $\mathfrak{R}'$ and $\mathfrak{R}''$. Now $V = V_1 \oplus V_2$ and $W = W_1 \oplus W_2$, where $W_2 = W \cap V_2$ $(= \sum_{i \in J} R_i W_1)$. Hence

$$V/W = V_1/W_1 \oplus V_2/W_2.$$

The representation of G induced by $\mathfrak{R}'$ is the proper part of the representation of G carried by W. Since W_1 carries the proper representation $\mathfrak{T}'$ of G, while W_2 carries a null representation of G (since $W_2 \subseteq V_2$), it follows that $\mathfrak{R}'$ induces $\mathfrak{T}'$. Similarly, V_1/W_1 carries $\mathfrak{T}''$ and V_2/W_2 carries a null representation of G, and so $\mathfrak{R}''$ induces $\mathfrak{T}''$ in G. Hence $\mathfrak{R}'$ is an extension of $\mathfrak{T}'$, and $\mathfrak{R}''$ an extension of $\mathfrak{T}''$ to S.

By an evident induction on the number r of irreducible constituents $\mathfrak{T}_i$ of $\mathfrak{T}$, it is clear that $\mathfrak{T}^*$ reduces into r representations $\mathfrak{R}_i$ such that $\mathfrak{R}_i$ is an extension of $\mathfrak{T}_i$ $(i = 1, \cdots, r)$. By Theorem 6.2 of [1] $\mathfrak{R}_i$ reduces into the basic extension $\mathfrak{T}_{i0}^*$ of $\mathfrak{T}_i$ and (possibly) null representations. By Theorem 7.1, each $\mathfrak{T}_{i0}^*$ is irreducible. Hence the non-null irreducible constituents of $\mathfrak{T}^*$ are precisely $\mathfrak{T}_{10}^*, \cdots, \mathfrak{T}_{r0}^*$.

The final assertion of the theorem is immediate from the foregoing and Theorem 7.1.

4. Examples. In Theorem 2, let $\mathfrak{T}^*$ be the basic extension $\mathfrak{T}_0^*$ of $\mathfrak{T}$. One might expect that each irreducible constituent of $\mathfrak{T}_0^*$ is the basic extension of one of the irreducible constituents of $\mathfrak{T}$. The following example shows that $\mathfrak{T}_0^*$ may have an extraneous null constituent.

Let G be the cyclic group $\{e, a\}$ of order 2. Let S be the Rees 2×2 matrix semigroup over G with "sandwich" matrix $P = \begin{pmatrix} e & e \\ e & a \end{pmatrix}$. Let Ω be the integers mod 2. Let

$$T(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad T(a) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

It is found that the basic extension $\mathfrak{T}_0^*$ of $\mathfrak{T}$ to S has degree 3 over Ω, and reduces into two unit representations and one null representation of S.

According to Theorem 1, a representation $\mathfrak{T}$ of G is extendible to S if its indecomposable constituents are extendible. The following example shows that the irreducible constituents of $\mathfrak{T}$ may be extendible, yet $\mathfrak{T}$ is not extendible.

Let G, Ω, and $\mathfrak{T}$ be as in the previous example. Let N be the set of natural numbers, and let S be the Rees $N \times N$ matrix semigroup over G with sandwich matrix $P = (p_{ij})$ given by

$$p_{ij} = \begin{cases} e & \text{if } i = 1 \text{ or } j = 1, \\ a & \text{otherwise.} \end{cases}$$

Since the irreducible constituents of $\mathfrak{T}$ are just unit representations of G, they are trivially extendible to S. But $\mathfrak{T}$ itself is found not to be extendible.

The Tulane University of Louisiana.

REFERENCES.

[1] A. H. Clifford, "Matrix representations of completely simple semigroups," *American Journal of Mathematics*, vol. 64 (1942), pp. 327-342.

[2] D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

[3] A. Suschkewitsch, "Über die Matrizendarstellung der verallgemeinerten Gruppen," *Communications de la Société Mathématique de Kharkow*, vol. 6 (1933), pp. 27-38.

SUR LA THÉORIE DE LA VARIÉTÉ DE PICARD.*

par C. CHEVALLEY.

En hommage amical et respecteux au professeur Zariski.

Introduction. Soit U une variété complète. On appelle habituellement diviseurs de U les combinaisons linéaires formelles d'hypersurfaces de U à coefficients entiers, c'est-à-dire les cycles de codimension 1 sur U. Cartier a introduit récemment une autre notion de diviseur (cf. [1]) ; si U est normale, ce que nous supposerons ici, un diviseur au sens de Cartier est un cycle de codimension 1 qui est localement principal, i. e. qui coincide au voisinage de chaque point avec le diviseur d'une fonction. Si U est non singulière, les deux notions de diviseur sont équivalentes, mais il n'en est plus de même en général.

Nous nous proposons ici d'étendre aux diviseurs de Cartier (que nous appellerons désormais simplement diviseurs) la notion de variété de Picard, tout au moins dans le cas des variétés complètes U qui sont normales. W. L. Chow a observé que, pour la notion classique de diviseurs, la variété de Picard d'une variété quelconque U était identique à la variété de Picard de la variété d'Albanese de U. Nous montrons que ce résultat reste vrai pour les diviseurs au sens de Cartier à condition de remplacer la variété d'Albanese de U par ce que nous appelons sa variété d'Albanese stricte: alors que la variété d'Albanese résoud le problème relatif aux fonctions (non partout définies) sur U à valeurs dans des variétés abéliennes, la variété d'Albanese stricte résoud le problème correspondant relatif aux morphismes (partout définis) de U dans des variétés abéliennes. Comme une variété abélienne est non singulière, il n'y a pas de différence entre diviseurs au sens classique et diviseurs de Cartier sur une telle variété; nous aurions donc pu tenir pour acquies la notion de variété de Picard d'un variété abélienne pour établir le résultat cité ci-dessus. Nous avons préféré reprendre la question dans son ensemble, car, d'une part, la démonstration que nous donnons de l'existence d'une variété de Picard pour une variété abélienne est, croyons-nous, plus simple que les démonstrations déjà connues, et d'autre part, elle se poursuit dans un esprit tout différent.

* Received June 14, 1959.

Nous utilisons systématiquement la notion de famille algébrique de diviseurs (ou de classes de diviseurs) d'une variété U paramétrée par une variété T: une famille algébrique de classes de diviseurs de U paramétrée par T est une application f de T dans le groupe des classes de diviseurs de U qui satisfait à certaines conditions. Supposons de plus que l'on ait $f(t_0) = 0$ pour un certain point t_0 de T. La famille f définit un diviseur de U (ou plutôt d'une variété déduite de U par extension du corps de base) rationnel sur le corps $F(T)$ de la variété T, donc un point de la variété de Picard P de U rationnel sur $F(T)$, c'est-à-dire une fonction sur T à valeurs dans P; nous montrons que (au moins si T est normale) cette fonction est partout définie.

Chapitre I. Familles de diviseurs.

I. Diviseurs. Nous utiliserons la définition de la notion de diviseur sur une variété dûe à Cartier ([1]). Rappelons qu'il y a une correspondence biunivoque entre les diviseurs sur une variété U et les faisceaux cohérents d'idéaux fractionnaires principaux sur U (quand nous parlerons d'idéaux fractionnaires, il sera toujours sous-entendu qu'il s'agit d'idéaux $\neq \{0\}$). Si d est un diviseur, nous désignerons par A^d le faisceau qui correspond à d; pour tout faisceau F de groupes sur U, nous désignerons par F_x le groupe ponctuel de F en un point x de U; $A^d{}_x$ est donc un idéal fractionnaire principal pour l'anneau local $\mathfrak{o}(x)$ de x; tout générateur de cet idéal est appelé une *fonction de définition de d en x.* Toute fonction sur U qui est fonction de definition de d en x l'est aussi en tous les points d'un voisinage de x. Les diviseurs pour lesquels il existe une fonction qui est une fonction de définition en tous les points de U sont les *diviseurs principaux.* Si les fonctions de définion de d en x sont définies en x, on dit que d est *positif en x*; si d est positif en tous les points de U, on dit qu'il est *positif,* et on écrit $d \geqq 0$. Les section du faisceau A^d sont les fonctions numériques v telles que $v \in A^d{}_x$ pour tout $x \in U$; ces fonctions sont aussi dites être des *multiples de d.*

Les opérations du calcul sur les idéaux fractionnaires définissent des opérations sur les faisceaux d'idéaux fractionnaires. Ainsi, si A et A' sont des faisceaux d'idéaux fractionnaires, les symboles AA', $A + A'$ représentent des faisceaux d'idéaux fractionnaires dont les idéaux ponctuels en un point x sont $A_xA'_x$ et $A_x + A'_x$ respectivement; si d et d' sont des diviseurs, on a $A^dA^{d'} = A^{d+d'}$. Soit O le faisceaux des anneaux locaux sur U; si A est un faisceau d'idéaux fractionnaires, le transporteur B de A dans O est un faisceau d'idéaux fractionnaires dont l'idéal ponctuel en un point x est l'ensemble des fonctions numériques v telles que $vA_x \subset O_x$. Si d est un diviseur, le transporteur de A^d dans O est A^{-d}.

A tout partie fermée $E \neq U$ de la variété U est associé un faisceau A^E d'idéaux fractionnaires de U; les sections de A^E sur un ouvert affine U' sont les fonctions partout définies sur U' qui sont nulles sur $U' \cap E$; si $x \in U$, $A^E{}_x$ se compose des fonctions définies en x et nulles en tous les points de E appartenant à un voisinage convenable de x. Le faisceau A^E s'appelle le *faisceau de définition de E.*

Soit A un faisceau d'idéaux fractionnaires sur une variété U; si O est le faisceau des anneaux locaux de U, l'ensemble des points $x \in U$ tels que $A_x \subset O_x$ est une partie ouverte non vide de U; soit E le complémentaire de cet ensemble, et soit A^E le faisceau de définition de E. Montrons qu'il y a un exposant $k > 0$ tel que $(A^E)^k A$ soit un faisceau d'idéaux entiers. Soit en effet U' un ouvert affine de U; les sections de A sur U' forment un idéal fractionnaire $\mathfrak{a}$ pour l'algèbre affine $\mathfrak{o}(U')$ de U'; soit $\mathfrak{b}$ l'ensemble des $u \in \mathfrak{o}(U')$ tels que $u\mathfrak{a} \subset \mathfrak{o}(U')$; c'est un idéal (entier) de $\mathfrak{o}(U')$. Si $x \in U'$, l'idéal $\mathfrak{b}\mathfrak{o}(x)$ engendré par $\mathfrak{b}$ dans l'anneau local $\mathfrak{o}(x)$ de x est l'ensemble des $u \in \mathfrak{o}(x)$ tels que $u\mathfrak{a}\mathfrak{o}(x) \subset \mathfrak{o}(x)$, ou encore tels que $uA_x \subset \mathfrak{o}(x)$ (A_x étant l'idéal ponctuel de A en x); on a donc $\mathfrak{b}\mathfrak{o}(x) = \mathfrak{o}(x)$ si $x \notin E$. L'ensemble des zéros (dans U') des fonctions de l'idéal $\mathfrak{b}$ est donc contenu dans E; il en résulte, en vertu du théorème des zéros de Hilbert, que, si $\mathfrak{a}(E)$ est l'idéal des fonctions de $\mathfrak{o}(U')$ nulle sur $U' \cap E$, il y a un $k(U') > 0$ tel que $(\mathfrak{a}(E'))^{k(U')} \subset \mathfrak{b}$. Ceci montre que $(A^E{}_x)^{k(U')} A_x$ est un idéal entier pour tout $x \in U'$. Il suffit alors de prendre pour k le plus grand des nombres $k(U')$ pour tous les ouverts U' d'un recouvrement ouvert affine fini de U.

Si U est une variété complète et F un faisceau cohérent sur U, les sections de F sur U forment un espace vectoriel de dimension finie sur K ([5]), d'où il résulte que, si A est un faisceau d'idéaux fractionnaires sur U, les fonctions multiples de A forment un espace vectoriel de dimension finie sur K. Nous dirons d'une manière général qu'une variété U est *semi-complète* si, pour tout faisceau A d'idéaux fractionnaires sur U, les fonctions multiples de A forment un espace vectoriel de dimension finie sur K. Une variété complète est évidemment semi-complète. Montrons que, *si U_1 est une variété complète et U une sous-variété ouverte normale de U_1 telle que* $\dim(U - U_1) \leq \dim U - 2$ (*si* $U_1 \neq U$), *U est semi-complète.* On peut supposer que U_1 est elle-même normale; en effet, il existe une variété normale complète U_2 et un morphisme birationnel f de U_2 sur U_1 tels que (U_2, f) soit un revêtement de U_1. Si $U' = f^{-1}(U)$, et si f' est la restriction de f à U', (U', f') est un revêtement de U; comme f' est birationnel et U normale, f' est un isomorphisme de U' sur U. Supposons donc que U_1 soit normale. Soit A un faisceau d'idéaux fractionnaires sur U. On vient de voir qu'il existe une partie

fermée $E \neq U$ de U et un exposant $k > 0$ tels que $(A^E)^k A$ soit un faisceau d'idéaux entiers. Nous désignerons par E_1 l'adhérence de E dans U_1 et par A^{E_1} le faisceau d'idéaux fractionnaires sur U_1 qui définit l'ensemble E_1; soit enfin A_1 le transporteur de $(A^{E_1})^k$ dans le faisceau des anneaux locaux de U_1. Il suffira d'établir que, si u est une fonction numérique sur U qui est multiple de A, la fonction u_1 sur U_1 qui prolonge u est multiple de A_1, c'est-à-dire que $u_1(A^{E_1}{}_x)^k$ est un idéal entier quel soit $x \in U_1$. Soit S_1 une hypersurface de U_1; montrons que l'ordre de u_1 le long de S_1 est $\geqq -k$ (nous supposons $u_1 \neq 0$). Puisque $\dim(U_1 - U) \leqq \dim U - 2$, S_1 rencontre U; il en résulte tout de suite qu'il existe un point $y \in S_1 \cap U$ tel que toute composante irréductible de E passant par y soit contenue dans S_1. Soit t une fonction numérique définie en y et qui engendre l'idéal premier maximal de l'anneau local de $S_1 \cap U$; il y a alors voisinage U' de U tel que t soit nulle en tout point de $U' \cap E$; il en résulte que $t^k \in (A^E{}_x)^k$, d'où $t^k A_x \in \mathfrak{o}(x)$ et par suite $t^k u \in \mathfrak{o}(x)$ puisque u est multiple de A; on en déduit que l'ordre de u_1 le long de S_1 est $\geqq -k$. Soit maintenant x un point quelconque de U_1; nous voulons montrer que, si $v_1, \cdots, v_k \in A^{E_1}{}_x$, $u_1 v_1 \cdots v_k$ est définie en x. Comme U_1 est normale, il suffit de montrer qu'aucune hypersurface S_1 de U_1 passant par x ne peut être variété de pôles de $u_1 v_1 \cdots v_k$.[1] Supposons d'abord que $S_1 \not\subset E_1$; il y a alors un point $y \in S_1 \cap U$ qui n'appartient pas à E; $A^E{}_x$ est alors l'anneau local de x, d'où il résulte que u est définie en x, donc que u_1 appartient à l'anneau local de S_1; il en est de même de chacune des fonctions v_i, ces fonctions étant définies en x; il en résulte que S_1 n'est pas variété de pôles de $u_1 v_1 \cdots v_k$. Supposons ensuite que $S_1 \subset E_1$; chacune des fonctions v_i, étant nulle sur tous les points de l'intersection avec E_1 d'un voisinage convenable de x, appartient à l'idéal premier maximal $\mathfrak{p}$ de l'anneau local de S_1. On a donc $v_1 \cdots v_k \in \mathfrak{p}^k$; comme l'ordre de u_1 le long de S_1 est $\geqq -k$, $u v_1 \cdots v_k$ appartient à l'anneau local de S_1, et S_1 n'est pas variété de pôles de cette fonction. Notre assertion est donc établie.

On notera qu'il résulte en particulier de là que, si U_1 est une variété complète et normale, l'ensemble U des points simples de U_1 est une variété semi-complète.

Si U est une variété semi-complète, toute fonction numérique u partout définie sur U est constante. En effet, chacune des puissances de u est une

[1] Nous appelons zéro d'une fonction numérique u sur une variété U tout point $x \in U$ tel que $(0, x)$ soit adhérent au graphe de u dans $K \times U$, et pôle de u (si $u \neq 0$) tout zéro de u^{-1}. Toute composante irréductible de l'ensemble des zéros (ou des pôles) d'une fonction numérique est une hypersurface ([2], proposition 10, chap. IV, § I). Si U est normale, tout point de U en lequel u n'est pas définie est un pôle de u ([2], proposition 2, chap. V, § I).

section du faisceau des anneaux locaux; il y a donc un entier $n > 0$ tel que $1, u, \cdots, u^n$ soient linéairement dépendantes sur K, ce qui montre que u est algébrique sur K, d'où $u \in K$.

Si d est un diviseur sur un variété U, l'ensemble des points x tels que $A^d{}_x$ soit distinct de l'anneau local de x est un ensemble fermé $\neq U$, qu'on appelle le *support* de d et qu'on note Supp d. Si u est une fonction de définition de d en tous les points d'une partie ouverte non vide U' de U, $U' \cap \mathrm{Supp}\, d$ se compose des points $x \in U'$ tels que u et u^{-1} ne soient pas tous deux définis en x. Si d et d' sont des diviseurs, on a $\mathrm{Supp}(d + d') \subset \mathrm{Supp}\, d \cup \mathrm{Supp}\, d'$, $\mathrm{Supp}(-d) = \mathrm{Supp}\, d$.

Soit f un morphisme d'une variété V dans une variété U, et soit d un diviseur sur U dont le support ne contienne pas l'ensemble $f(V)$. Soient y un point de V et u une fonction de définition de d en $x = f(y)$. La fonction u est alors composable avec f, et on a $u \odot f \neq 0$. Soit en effet U' un voisinage ouvert de x tel que u soit fonction de définition de d en tout point de U'. Comme $f(V)$ est un ensemble irréductible non contenu dans $U' \cap \mathrm{Supp}\, d$, $U' \cap f(V)$ n'est pas contenu dans $U' \cap \mathrm{Supp}\, d$, ce qui montre qu'il y a un point $x' \in U' \cap f(V)$ tel que u et u^{-1} soient définies en x', ce qui établit notre assertion. Soit $\mathfrak{o}(y)$ l'anneau local de y sur V; on vérifie immédiatement que l'idéal fractionnaire $(u \odot f)\mathfrak{o}(y)$ ne dépend pas de choix de la fonction de définition u de d en x; soit B_y cet idéal fractionnaire. Il est clair que les B_y sont les idéaux ponctuels d'un faisceau d'idéaux fractionnaires principaux sur V; ce dernier définit un diviseur sur V, qu'on appelle l'*image réciproque de d par f*, et qu'on note $f^*(d)$. Les diviseurs d dont les supports ne contiennent pas $f(V)$ forment un sous-groupe du groupe des diviseurs de U, et f^* est un homomorphisme de ce groupe dans le groupe des diviseurs de V. En particulier, si V est une sous-variété de U, f étant l'injection canonique de V dans U, si $f^*(d)$ est défini, on dit que ce diviseur est le *diviseur induit par d sur V*. Si f est un morphisme dominant d'une variété V dans U (i.e. si $f(V)$ est dense dans U), $f^*(d)$ est défini quel que soit le diviseur d. En particulier, si V est une sous-variété ouverte d'une variété U, tout diviseur sur U induit un diviseur sur V. Mais il est important de remarquer que l'application ainsi définie du groupe des diviseurs sur U dans le groupe des diviseurs sur V n'est en général pas surjective; nous verrons cependant qu'elle l'est si U est non singulière.

Soient par ailleurs T une variété et p la projection du produit $T \times U$ sur son second facteur; comme p est surjectif, l'image réciproque par p de tout diviseur d sur U est définie; cette image réciproque sera souvent notée $T \times d$. On a $\mathrm{Supp}(T \times d) = T \times \mathrm{Supp}\, d$, comme il résulte du fait que, si u est

une fonction numérique sur U, $u \odot p$ n'est définie en un point $(t, x) \in T \times U$ que si u est définie en x. Désignons de plus par t_0 un point de T, par x_0 un point de U, par j l'application $x \to (t_0, x)$ de U dans $T \times U$ et par k l'application $t \to (t, x_0)$ de T dans $T \times U$. Alors $j^*(T \times d)$ est toujours défini et égal à d; $k^*(T \times d)$ n'est défini que si x_0 n'appartient pas à Supp d, et est alors nul. Ces faits résultent immédiatement des définitions.

Soient f un morphisme d'une variété V dans une variété U et g un morphisme d'une variété W dans V; soit d un diviseur sur U. Si $(f \circ g)^*(d)$ est défini, il en est de même de $f^*(d)$ et de $g^*(f^*(d))$, et on a

$$g^*(f^*(d)) = (f \circ g)^*(d).$$

Soit d un diviseur sur une variété *normale* U, et soit S une hypersurface de U. On sait que l'anneau local $\mathfrak{o}(S)$ de S sur U est un anneau local principal. Si $\mathfrak{p}_S$ est l'idéal premier maximal de cet anneau, les idéaux fractionnaires pour $\mathfrak{o}(S)$ sont les puissances d'exposants de signes quelconques de $\mathfrak{p}_S$. En particulier, on a, si $x \in S$, $A^d{}_x\mathfrak{o}(S) = \mathfrak{p}_S{}^{k(S)}$, $k(S)$ étant un entier dont on voit tout de suite qu'il ne dépend que de d et de S, non du choix de x sur S. Il est clair que $k(S) = 0$ si S n'est pas contenu dans Supp d; il n'y a donc qu'un nombre fini d'hypersurfaces S pour lesquelles $k(S) \neq 0$, et on peut associer à d le cycle $Z(d) = \sum_S k(S) S$ de codimension 1 sur U. On obtient ainsi un homomorphisme $d \to Z(d)$ du groupe des diviseurs dans le groupe des cycles de codimension 1. Nous allons montrer que cet homomorphisme est *injectif*. Il suffira pour cela d'établir que Supp d *est la réunion des hypersurfaces* S *pour lesquelles* $k(S) \neq 0$. On sait déjà que ces hypersurfaces sont contenues dans Supp d. Soient x un point de Supp d, et u une fonction de définition de d en x. L'une au moins des fonctions u, u^{-1} n'est pas définie en x. Comme U est normale, il en résulte que x est un zéro ou un pôle de u. Si par exemple x est un pôle de u, il passe par x une composante irréductible S de l'ensemble des pôles de u, et on sait que S est une hypersurface; comme u n'est définie en aucun point de S, on a $k(S) < 0$. On voit de même que, si x est un zéro de u, il passe par x une hypersurface S pour laquelle $k(S) > 0$.

On notera que le raisonnement qu'on vient de faire prouve que, si d n'est pas positif en un point x, il passe par x au moins une hypersurface S pour laquelle $k(S) < 0$.

Il est important de remarquer que l'application $d \to Z(d)$ n'est en général pas surjective; par exemple, si U est un cône quadratique, le cycle constitué par une génératrice du cône, prise avec le coefficient 1, n'est le cycle associé à aucun diviseur. On a cependant le résultat suivant:

Proposition 1. *Si U est une variété non singulière, tout cycle de codimension 1 sur U est associé à un diviseur sur U.*

Il suffit de montrer que, si S est une hypersurface, le cycle $1 \cdot S$ est associé à un diviseur. Soit A^S le faisceau d'idéaux qui définit S; comme U est non singulière, il est connu que, pour tout $x \in S$, l'idéal de définition de S en x est principal; A^S est donc un faisceau d'idéaux principaux, et est par suite associé à un diviseur d; il est clair que $Z(d) = 1 \cdot S$.

Corollaire. *Soient U une variété non singulière, V une sous-variété ouverte de U et i l'injection canonique de V dans U; i^* est alors une application surjective du groupe des diviseurs de U sur celui de V.*

En effet, si S_V est une hypersurface de V, son adhérence S dans U est une hypersurface de U; si d est le diviseur sur U auquel S est associée, le cycle sur V associé à $i^*(d)$ est S_V.

Proposition 2. *Soit h un morphisme surjectif propre d'une variété X' dans une variété X; supposons que, pour tout $x' \in X'$, toute fonction numérique u sur X telle que $u \odot h$ soit définie en x' soit définie au point $h(x')$. Soit d' un diviseur sur X'; supposons que, pour tout $x \in X$, il existe une fonction numérique v sur X telle que $v \odot h$ soit fonction de définition de d' en tout point de $h^{-1}(x)$. Il existe alors un diviseur d et un seul sur X tel que $d' = h^*(d)$; on a* $\operatorname{Supp} d' = h^{-1}(\operatorname{Supp} d)$.

Si $x \in X$, soit $\mathfrak{o}(x)$ l'anneau local de x. Soit v une fonction numérique sur X tell que $h \odot v$ soit fonction de définition de d' en tout point de $h^{-1}(x)$; l'idéal fractionnaire $\mathfrak{o}(x)v$ ne dépend alors que de x et d'; en effet, si v_1 est une autre fonction qui possède la même propriété que v, $h \odot (v^{-1}v\ )$ est définie en tout point de $h^{-1}(x)$ et y prend une valeur $\neq 0$, ce qui implique que $v^{-1}v_1$ est définie en x et y prend une valeur $\neq 0$, donc que $\mathfrak{o}(x)v = \mathfrak{o}(x)v_1$; posons $A_x = \mathfrak{o}(x)v$. Montrons qu'il y a un voisinage de x tel que l'on ait $A_y = \mathfrak{o}(y)v$ pour tout point de ce voisinage. L'ensemble X'_0 des points $x' \in X'$ tels que $h \odot v$ soit fonction de définition de d' en x' est ouvert; comme h est propre, $h(X' - X'_0)$ est fermé; de plus, cet ensemble ne contient pas x; $X_0 = X - h(X' - X'_0)$ est donc un voisinage de x, et il est clair que $A_y = \mathfrak{o}(y)v$ pour tout $y \in X_0$. Il résulte de là que les $A_x (x \in X)$ sont les idéaux ponctuels d'un faisceau d'idéaux fractionnaires principaux sur X; ce faisceau définit un diviseur d; il est clair que $d' = h^*(d)$ et que d est le seul diviseur sur X possédant cette propriété. Les notations étant comme ci-dessus, supposons que $x \in \operatorname{Supp} d$; alors l'une au moins des fonctions v, v^{-1} n'est pas définie en x. Il en résulte que, pour tout $x' \in h^{-1}(x)$, l'une au moins

des fonctions $h \odot v$, $(h \odot v)^{-1}$ n'est pas définie en x', ce qui montre que $h^{-1}(x) \subset \operatorname{Supp} d'$. Comme on sait déjà que $\operatorname{Supp} d'$ est contenu dans $h^{-1}(\operatorname{Supp} d)$, on a $\operatorname{Supp} d' = h^{-1}(\operatorname{Supp} d)$.

Remarque. Supposons que (X', h) soit un revêtement de X. La condition que nous avons imposée à h dans l'énoncé de la Proposition 2 est alors satisfaite dans le cas où X est normale ([2], chap. V, § V, proposition 4). Elle l'est également si on suppose que h est un revêtement galoisien non ramifié ([3], chap. V, § 11, Corollaire 2 à la Proposition 7); rappelons que cela signifie qu'il existe un groupe G d'automorphismes de X' tel que les orbites relativement à G des points de X' soient exactement les ensembles $h^{-1}(x)$, $x \in X$, et que de plus h n'est ramifié en aucun point de $h^{-1}(x)$, ce qui peut se traduire par la condition que G opère sans point fixe sur X'.

Nous utiliserons dans la suite le lemme suivant:

LEMME 1. *Soient d un diviseur sur une variété U et $x_1, \cdots, x_m$ un nombre fini de points de U qui appartiennent à un même morceau affine de U; il y a alors une fonction numérique sur U qui est fonction de définition de d en chacun des points x_i.*

On peut supposer les points x_i mutuellement distincts. Soit u_i une fonction de définition de d en x_i; on a donc $\operatorname{div} u_i = d + d_i$, où d_i est un diviseur dont le support ne contient pas x_i (nous notons $\operatorname{div} u$ le diviseur principal associé à une fonction $u \neq 0$). Soit J_i le faisceau d'idéaux associé à l'ensemble $\operatorname{Supp} d_i$; il résulte de ce qui a été dit plus haut qu'il existe un entier $k_i \geqq 0$ tel que $J_i^{k_i} A^{d_i}$ soit un faisceau d'idéaux entiers. Soit U' un morceau affine de U contenant les points x_i; comme $x_i \notin \operatorname{Supp} d_i$, il y a une fonction numérique z_i partout définie sur U' qui est nulle sur $U' \cap \operatorname{Supp} d_i$, qui prend la valeur 0 en tous les points x_j, $j \neq i$, mais qui ne prend pas la valeur 0 en x_i. Posons $u_i' = u_i z_i^{k_i+1}$; u_i' est encore fonction de définition de d en x_i; si $\operatorname{div} u_i' = d + d_i'$, d_i' est la somme de $\operatorname{div} z_i$ et d'un diviseur qui est $\geqq 0$ en tout point de U', ce qui montre que $x_j \in \operatorname{Supp} d_i'$ si $j \neq i$; par contre, x_i n'est pas dans $\operatorname{Supp} d_i'$. Soit $u = \sum_{i=1}^{m} u_i'$; on a alors $u u_i'^{-1} = 1 + \sum_{j \neq i} u_j' u_i'^{-1}$; si $j \neq i$, on a $\operatorname{div} u_j' u_i'^{-1} = d_j' - d_i'$; or d_j' est positif en x_i et $x_i \notin \operatorname{Supp} d_i'$; il en résulte que $u_j' u_i'^{-1}$ est définie en x_i; comme $x_i \in \operatorname{Supp} d_j'$, on a $(u_j' u_i'^{-1})(x_i) = 0$. Il résulte de là que $u u_i'^{-1}$ est définie en x_i et y prend la valeur 1, donc que u est fonction de définition de d en x_i. Ceci étant vrai pour tout i, le lemme est établi.

PROPOSITION 3. *Soit (X', h) un revêtement galoisien non ramifié d'une*

variété X; supposons que, pour tout $x \in X$, $h^{-1}(x)$ soit contenu dans un morceau affine de X'. Soit d' un diviseur sur X' tel que l'on ait $s^*(d') = d'$ pour tout automorphisme s du revêtement (X', h); il existe alors un diviseur d et un seul sur X tel que $d' = h^*(d)$; on a Supp $d' = h^{-1}$(Supp d).

Tenant compte de la Proposition 2 et de la remarque qui suit la démonstartion de cette proposition,[2] on voit qu'il suffit de montrer que, si $x \in X$, il y a une fonction numérique v sur X telle que $v \odot h$ soit fonction de définition de d' en tout point de $h^{-1}(x)$. Il existe une fonction numérique v' sur X' qui est fonction de définition de d' en tout point de $h^{-1}(x)$ (Lemme 1). Il est clair que, si G est le groupe des automorphismes du revêtement (X', h), toute fonction de la forme $v' \odot s$ $(s \in G)$ possède la même propriété que v'; si donc $x' \in h^{-1}(x)$, $v'^{-1}(v' \odot s)$ est définie et prend une valeur $a_s \neq 0$ en x'. Puisque h est non ramifié, on a $s(x') \neq x'$ pour toute opération s distincte de l'identité de G. Comme $h^{-1}(x)$ est contenu dans un morceau affine de X', il existe une fonction numérique z' sur X', définie en tout point de $h^{-1}(x)$, telle que $z'(x') = 1$, $z'(s(x')) = 0$ pour tout $s \in G$ distinct de l'identité. Posons $v'_1 = \sum_{s \in G} (z'v' \odot s)$; $v'^{-1}v'_1$ est définie en x' et y prend la valeur $a_e \neq 0$, ce qui signifie que v'_1 est fonction de définition de d' en x'. Or on a $v' \odot s = v'_1$ pour tout $s \in G$; il en résulte d'abord que v'_1 est fonction de définition de d' en tout point de $h^{-1}(x)$, puis (le revêtement h étant galoisien, donc séperable) que v'_1 se met sous la form $v \odot h$, v étant une fonction numérique sur X. La Proposition 3 est donc étable.

Remarque. La condition que, pour tout $x \in X$, $h^{-1}(x)$ soit contenu dans un morceau affine de X' est satisfaite si X' est normale; soit alors en effet X_0 un morecau affine de X contenant x; posons $X'_0 = h^{-1}(X_0)$, et désignons par h_0 la restriction de h à X'_0; (X'_0, h_0) est alors un revêtement normal de la variété affine X_0, d'où il résulte que X'_0 est affine. On peut montrer que la condition en question est satisfaite pour tout revêtement; mais nous n'aurons pas besoin de ce résultat plus fin.

II. Familles algebriques de diviseurs. Soient T et U des variétés. Pour tout $t \in T$, nous désignerons par j_t l'application $x \to (t, x)$ de U dans $T \times U$.

DÉFINITION 1. *Une application f de T dans le groupe des diviseurs de U s'appelle une famille algébrique de diviseurs de U (paramétrée par T)*

[2] Il suffira d'ailleurs pour la suite de savoir que la Proposition 3 est vraie dans le cas où X est normale.

s'il existe un diviseur D de $T \times U$ tel que, pour tout $t \in T$, $j_t^*(D)$ soit défini et égal à $f(t)$. On dit alors que D est un diviseur de définition de la famille f.

Il est clair que les familles de diviseurs de U paramétrées par T forment un groupe additif.

PROPOSITION 1. *Soit f une famille algébrique de diviseurs de U paramétrée par T; soit h un morphisme d'une variété T' dans la variété T. Alors $f \circ h$ est une famille algébrique de diviseurs de U paramétrée par T'; si D est un diviseur de définition de $f, h^*(D)$ est défini et est un diviseur de définition de $f \circ h$.*

Dans l'énoncé précédent, ainsi qu'en plusieurs endroits de la suite de ce mémoire, nous faisons la convention de notation suivante: si h est un morphisme de T' dans T, nous désignons encore par h le morphisme $(t', x) \rightarrow (h(t'), x)$ de $T' \times U$ dans $T \times U$.

Si $t' \in T'$, soit $j'_{t'}$ l'application $x \rightarrow (t', x)$ de U dans $T' \times U$; on a $j_{h(t')} = h \circ j'_{t'}$, et $j_{h(t')}^*(D)$ est défini et égal à $f(h(t'))$; on en conclut que $h^*(D)$ est défini, qu'il en est de même de $(j'_{t'})^*(h^*(D))$, et que ce dernier diviseur est égal à $j_{h(t')}^*(D) = f(h(t'))$, ce qui démontre la Proposition 1.

PROPOSITION 2. *Soit f une famille algébrique de diviseurs de U paramétrée par T; soit g un morphisme dominant d'une variété U' dans U. Alors $t \rightarrow g^*(f(t))$ est une famille algébrique de diviseurs de U'; si D est un diviseur de définition de $f, g^*(D)$ est un diviseur de définition de la famille $t \rightarrow g^*(f(t))$.*

On notera que $g^*(D)$ est défini puisque l'application $(t, x') \rightarrow (t, g(x'))$ est un morphisme dominant de $T \times U'$ dans $T \times U$. Soit t un point de T; il existe un point $x \in U$ tel que $(t, x) \notin \operatorname{Supp} D$; de plus, les points qui possèdent cette propriété forment une partie ouverte de U, qui rencontre donc $g(U')$; il y a donc un point $x' \in U'$ tel que $(t, g(x')) \notin \operatorname{Supp} D$, d'où $(t, x') \notin \operatorname{Supp} g^*(D)$. Il en résulte que, si on désigne par j'_t l'application $x' \rightarrow (t, x')$ de U' dans $T \times U'$, $(j'_{t'})^*(g^*(D))$ est défini; de plus, $(g \circ j'_t)^*(D)$ est défini. Comme $g \circ j'_t = j_t \circ g$, il en résulte que

$$(j'_t)^*(g^*(D)) = (g \circ j'_t)^*(D) = g^*(j_t^*(D)) = g^*(f(t)),$$

ce qui démontre la Proposition 2.

THÉORÈME 1. *Soient T et U des variétés et f une famille algébrique de diviseurs de U paramétrée par T; soit D un diviseur de définition de f. Si on a $f(t) \geqq 0$* (resp. $f(t) = 0$) *pour tous les points t d'une partie dense de T,*

on a $D \geqq 0$ (resp. $D = 0$), *et par suite* $f(t) \geqq 0$ (resp. $f(t) = 0$) *pour tous les points de* T.

Les assertions relatives au cas où $f(t) = 0$ pour tous les points d'une partie dense de T se déduisent de celles relatives au cas où $f(t) \geqq 0$ pour tous les points d'une partie dense en observant qu'une condition nécessaire et suffisante pour qu'un diviseur d sur une variété soit nul est que l'on ait à la fois $d \geqq 0$ et $-d \geqq 0$. Il nous suffira donc de prouver les premières de ces assertions. Montrons d'abord qu'on peut se ramener au cas où T et U sont affines. Soit (t_0, x_0) un point de $T \times U$; on veut montrer que D est positif en ce point. Soient T_0 et U_0 des morceaux affines de T et U contenant t_0 et x_0 respectivement; soit i l'application canonique de U_0 dans U. La restriction de l'application $t \to i^*(f(t))$ à T_0 est la famille de diviseurs de U_0 définie par le diviseur D_0 induit par D sur $T_0 \times U_0$. L'ensemble des points $t \in T_0$ tels que $i^*(f(t)) \geqq 0$ est dense; si le théorème est prouvé pour les variétés affines, il en résultera que $D_0 \geqq 0$, donc que D est positif en (t_0, x_0).

Supposant T et U affines, montrons qu'on peut se ramener au cas où D est principal. Soit (t_0, x_0) un point de $T \times U$, et soit w une fonction de définition de D en ce point; on a donc $\operatorname{div} w = D + D'$, où D' est un diviseur dont le support ne contient pas (t_0, x_0). Puisque $T \times U$ est une variété affine, il y a une fonction polynome z sur cette variété qui est nulle sur $\operatorname{Supp} D'$ mais qui prend une valeur $\neq 0$ en (t_0, x_0). On voit alors facilement qu'il y a un exposant $k \geqq 0$ tel que $\operatorname{div} z^k + D' \geqq 0$ (cf. § I). Soit $w' = wz^k$; w' est encore une fonction de définition de D en (t_0, x_0), et on a $\operatorname{div} w' = D + D''$, où D'' est un diviseur $\geqq 0$. Il y a un voisinage affine T_1 de t_0 dans T tel que $T_1 \times \{x_0\}$ ne rencontre pas $\operatorname{Supp} D''$; si $t \in T_1$, il résulte du fait que $j_t^*(D) = f(t)$ est défini qu'il en est de même de $j_t^*(\operatorname{div} w)$; de plus, $j_t^*(\operatorname{div} w) = f(t) + j_t^*(D'')$, d'où il résulte que $j_t^*(\operatorname{div} w)$ est $\geqq 0$ pour tous les points d'une partie dense de T_1; si on peut en conclure que $\operatorname{div} w \geqq 0$, il en résultera que D est positif en (t_0, x_0).

Supposons à partir de maintenant que T et U soient affines et que $D = \operatorname{div} w$ soit principal. Nous allons montrer qu'on peut se ramener au cas où on suppose de plus que T et U sont normales. Supposons le théorème établi dans ce cas. Il existe des variétés normales T' et U' et des morphismes $h: T' \to T$, $g: U' \to U$ tels que h et g soient des morphismes de revêtement; il est bien connu que T' et U' sont alors encore des variétés affines. Soit r le morphisme $(t', x') \to (h(t'), g(x'))$ de $T' \times U'$ dans $T \times U$; il résulte des Propositions 1 et 2 que l'application $f': t' \to g^*(h(t'))$ est une famille algébrique de diviseurs de U' paramétrée par T' qui admet $r^*(D)$ comme diviseur

de définition. Si A est l'ensemble des points $t \in T$ tels que $f(t) \geqq 0$, on a $f'(t') \geqq 0$ toutes les fois que $t' \in h^{-1}(A)$; comme A est dense dans T, $h^{-1}(A)$ est dense dans T'. En vertu de l'hypothèse faite, il en résulte que le diviseur principal $r^*(D)$ est $\geqq 0$. Si $D = \operatorname{div} w$, on a $r^*(D) = \operatorname{div}(w \odot r)$; $w \odot r$ est donc une fonction numérique partout définie sur $T' \times U'$. Observons par ailleurs que, si $t \in A$, $f(t) = j_t^*(\operatorname{div} w) = \operatorname{div} w \odot j_t$ est partout définie sur U. On va montrer que l'espace vectoriel V engendré par les fonctions $w \odot j_t$ pour tous les $t \in A$ est de dimension finie. Soit t un point de A, et soit t' un point de $h^{-1}(t)$. On a $j_t \circ g = r \circ j'_{t'}$, où $j'_{t'}$ est l'application $x' \rightarrow (t', x')$ de U' dans $T' \times U'$; il en résulte que $w \odot j_t \odot g = (w \odot r) \odot j'_{t'}$. Par ailleurs, l'application $u \rightarrow u \odot g$ est un isomorphisme du corps des fonctions numériques sur U sur un sous-corps du corps des fonctions numériques sur U'. Pour montrer que V est de dimension finie, il suffira donc de montrer que l'espace vectoriel engendré par les fonctions $(w \odot r) \odot j'_{t'}$ (pour tous les points $t' \in T'$) est de dimension finie dans le corps des fonctions sur U'. Mais, comme $w \odot r$ est partout définie sur $T' \times U'$, il y a des fonctions θ'_i $(1 \leqq i \leqq h)$ partout définies sur T' et u'_i partout définies sur U' telles que $(w \odot r)(t', x') = \sum_{i=1}^{h} \theta'_i(t') u'_i(x')$ pour tout $(t', x') \in T' \times U'$; $(w \odot r) \odot j'_{t'}$ est donc toujours une combinaison linéaire de $u'_1, \cdots, u'_h$. Ceci étant, soit $(u_1, \cdots, u_m)$ une base de V; si $t \in A$, il y a des éléments $\theta_i(t)$ $(1 \leqq i \leqq m)$ de K tels que $w \odot j_t = \sum_{i=1}^{m} \theta_i(t) u_i$. Il s'agit de montrer que les applications θ_i peuvent se prolonger en des fonctions numériques partout définies sur T. Soit t_1 un point de T; comme $j_{t_1}^*(\operatorname{div} w)$ est défini, l'ensemble U_1 des points $x \in U$ tels que w soit définie en (t_1, x) est ouvert et non vide. Comme $u_1, \cdots, u_m$ sont linéairement indépendantes, il est facile de voir qu'il existe des points $x_1, \cdots, x_m$ de U_1 tels que $\det(u_i(x_j)) \neq 0$. Comme w est définie aux points (t_1, x_j), il y a un voisinage ouvert T_1 de t_1 tel que w soit définie en tout point de chacun des ensembles $T_1 \times \{x_j\}$. Soit t un point de $T_1 \cap A$; alors les $\theta_i(t)$ peuvent s'obtenir par la résolution du système d'équations linéaires $\sum_{i=1}^{m} \theta_i(t) u_i(x_j) = w(t, x_j)$ $(1 \leqq j \leqq m)$. Or les application $t \rightarrow w(t, x_j)$ $(t \in T)$ sont des fonctions numériques partout définies sur T_1; il en résulte aussitôt que les restrictions des θ_i à $A \cap T_1$ peuvent se prolonger en des fonctions numériques partout définies sur T_1. Soient maintenant T_1 et T_2 des parties ouvertes non vides de T tells que les restrictions des θ_i à $T_1 \cap A$ (resp. $T_2 \cap A$) puissent se prolonger en des fonctions numériques $\theta_{i;1}$ (resp. $\theta_{i;2}$) sur T_1 (resp. T_2). Alors, pour chaque i, $\theta_{i;1}$ coincide avec $\theta_{i;2}$ sur

l'ensemble $A \cap T_1 \cap T_2$, qui est dense dans $T_1 \cap T_2$; il en résulte que $\theta_{i;1}$ coincide avev $\theta_{i;2}$ sur $T_1 \cap T_2$. Il résulte tout de suite de là que les applications θ_i peuvent, d'une manière et d'une seule, se prolonger en des fonctions numériques partout définies sur T, que nous désignerons encore par θ_i. Soit alors w_0 la fonction numérique partout définie sur $T \times U$ donnée par la formule $w_0(t,x) = \sum_{i=1}^{m} \theta_i(t) u_i(x)$; si W est l'ensemble de définition de w, $w - w_0$ prend la valeur 0 en tout point de l'ensemble $W \cap (A \times U)$, qui est dense dans W; elle est donc nulle, ce qui montre que $w = w_0$, donc que w est partout définie et par suite que $\operatorname{div} w \geqq 0$.

Il nous reste à démontrer le théorème dans le cas où T et U sont normales et où $D = \operatorname{div} w$ est un diviseur principal. Pour montrer que w est partout définie sur $T \times U$, il suffira, puisque $T \times U$ est normale, de montrer que w n'a pas de pôle. Or, il est bien connu que, si w avait au moins une variété de pôles, soit S, il existerait un point $(t,x) \in S$ tel que w^{-1} soit définie et prenne la valuer 0 en (t,x); mais alors la fonction $(w \odot j_t)^{-1}$ serait définie et prendrait la valuer 0 en x, de sorte que x serait un pôle de $w \odot j_t$ et que $f(t) - \operatorname{div} w \odot j_t$ ne serait pas un diviseur $\geqq 0$. Le Théorème 1 est donc établi.

Corollaire 1. *Si f est une famille algébrique de diviseurs d'une variété U paramétrée par une variété T, il n'y a qu'un seul diviseur sur $T \times U$ qui soit diviseur de définition de la famille f.*

Cela résulte immédiatement du Théorème 1.

Corollaire 2. *Soit f une application d'une variété T dans l'ensemble des diviseurs d'une variété U. Supposons que chaque point de T ait un voisinage ouvert T' tel que la restriction de f à T' soit une famille algébrique paramétrée par T'. L'application f est alors une famille algébrique de diviseurs.*

Il existe un recouvrement $(T_i)_{i \in I}$ de T par des ensembles ouverts non vides tels que, pour tout i, la restriction de f à T_i soit une famille algébrique paramétrée par T_i; soit D_i le diviseur de définition de cette famille; c'est un diviseur de $T_i \times U$. Si $i, j \in I$, les diviseurs induits par D_i et D_j sur $(T_i \cap T_j) \times U$ définissent la même famille algébrique de diviseurs, et sont par suite égaux. Il en résulte immédiatement qu'il existe un diviseur D sur $T \times U$ tel que, pour tout i, D_i soit le diviseur induit par D sur $T_i \times U$. Il est clair que l'on a $j_t^*(D) = f(t)$ pour tout $t \in T$, ce qui démontre le Corollaire 2.

Corollaire 3. *Soit f une famille algébrique de diviseurs d'une variété*

U paramétrée par une variété T. L'ensemble des points $t \in T$ tels que $f(t) \geqq 0$ est fermé, et il en est de même de l'ensemble des points t tels que $f(t) = 0$.

Comme dans la démonstration du Théorème 1, il suffit de démontrer que l'ensemble E des points t tels que $f(t) \geqq 0$ est fermé. Soient E_1 une composante irréductible de E et T' son adhérence; la restriction de f à T' est une famille algébrique paramétrée par T' et qui fait correspondre des diviseurs $\geqq 0$ aux points de la partie dense E_1 de T'; elle est donc positive, d'où $T' \subset E$. Ceci étant vrai pour toute composante irréductible de E, E est fermé.

COROLLAIRE 4. *Soient f et f' des familles algébriques de diviseurs d'une même variété U paramétrées par des variétés T et T'. L'ensemble des points $(t, t') \in T \times T'$ tels que $f(t) = f'(t')$ est alors fermé.*

Cela résulte du Corollaire 3 et du fait que l'application $(t, t') \rightarrow f(t) - f'(t')$ est une famille algébrique paramétrée par $T \times T'$ (Proposition 1).

Nous allons maintenant donner deux exemples importants de familles algébriques de diviseurs.

Soient U une variété et V un sous-espace vectoriel de dimension finie > 0 du corps $F(U)$ des fonctions numériques sur U. Puisque V est de dimension finie, l'ensemble U_0 des points de U en lesquels toutes les fonctions de V sont définies est ouvert et non vide. L'application $(u, x) \rightarrow u(x)$ de $V \times U_0$ dans K se prolonge en une fonction numérique w sur $V \times U$. Soit V_0 l'ensemble des éléments $\neq 0$ de V; si $u \in V$, soit j_u l'application $x \rightarrow (u, x)$ de U dans $V \times U$; il est clair que w est toujours composable avec j_u, et que $w \odot j_u = u \neq 0$ si $u \in V_0$. Soit w_0 la restriction de w à $V_0 \times U$; il résulte de ce qu'on vient de dire que $\operatorname{div} w_0$ est le diviseur de définition d'une famille algébrique de diviseurs de U paramétrée par V_0, qui n'est autre que l'application $u \rightarrow \operatorname{div} u (u \in V_0)$.

Désignons maintenant par $\mathfrak{P}(V)$ l'espace projectif associé à V, qui se compose des sous-espace de dimension 1 de V, et par φ l'application canonique $u \rightarrow Ku$ de V_0 sur $\mathfrak{P}(V)$; comme $\operatorname{div} cu = \operatorname{div} u$ si c est un élément $\neq 0$ de K, on voit que $\operatorname{div} u$ ne dépend que du point $\zeta = \varphi(u)$. Nous poserons $\operatorname{div} u = \operatorname{div} \zeta$ si $\zeta = \varphi(u)$. Montrons que l'application $\zeta \rightarrow \operatorname{div} \zeta$ est une famille algébrique de diviseurs de U paramétrée par $\mathfrak{P}(V)$. Soit ζ_0 un point de $\mathfrak{P}(V)$. Il existe alors un voisinage ouvert T_0 de ζ_0 dans $\mathfrak{P}(V)$ et un morphisme r de T_0 dans V_0 tels que $\varphi \circ r$ soit l'application indentique de T_0; si $\zeta \in T_0$, on a $\operatorname{div} \zeta = \operatorname{div} r(\zeta)$, d'où il résulte que la restriction à T_0 de l'application $\zeta \rightarrow \operatorname{div} \zeta$ est une famille algébrique de diviseurs de T_0; on conclut alors au moyen du Corollaire 2 au Théorème 1. L'application $\zeta \rightarrow \operatorname{div} \zeta$ s'appelle le *système linéaire de diviseurs de U défini par V.*[8] Si U est une variété semi-

complète, l'application $\zeta \to \operatorname{div} \zeta$ est injective, car toute fonction numérique de diviseur nul sur U est alors constante.

Soit maintenant C une courbe normale. Soit r un entier $\geqq 0$; soit C^r le produit de r exemplaires de C. Les permutations des facteurs du produit C^r définissent un groupe fini P d'automorphismes de C^r. Il est bien connu que C est une variété quasi-projective; il en est donc de même de C^r, de sorte que toute partie affine de C^r est contenue dans un morceau affine de cette variété. Il existe donc une variété quotient S^r de C^r par P: il existe un morphisme s_r de C^r sur S^r tel que (C^r, s_r) soit un revêtement galoisien de groupe P de S^r, et S^r est normale. On dit que S^r est la *puissance symétrique* r-ième de C, et s_r le *morphisme canonique* $C^r \to S^r$.

Soit maintenant f une famille algébrique de diviseurs d'une variété normale U paramétrée par C. Il résulte immédiatement de la Proposition 1 que l'application

$$(x_1, \cdots, x_r) \to \sum_{i=1}^{r} f(x_i)$$

est une famille algébrique m de diviseurs de U paramétrée par C^r. Nous allons montrer que m peut se mettre sous la fome $g \circ s_r$, où g est une famille algébrique de diviseurs de U paramétrée par S_r. Soit M le diviseur de définition de m. Soient $(a_1, \cdots, a_r)$ un point de C^r, b un point de U et w une fonction numérique sur $C \times U$ qui est fonction de définition de M en chacun des points (a_i, b) (Lemme 1, § I; on notera que, C étant une variété quasi-projective, toute partie finie de C est contenue dans un morceau affine). Soient $q_1, \cdots, q_r$ les projections de C^r sur ses divers facteurs; nous désignons encore par q_i l'application $((x_1, \cdots, x_r), y) \to (x_i, y)$ de $C^r \times U$ sur $C \times U$. Il est clair que $\prod_{i=1}^{r} (w \odot q_i)$ est une fonction de définition de M en $((a_1, \cdots, a_r), b)$; soit z cette fonction. Il est clair que, pour toute opération p du groupe P, on a $z \odot p = z$ (p étant identifié à l'application $((x_1, \cdots, x_r), y) \to (p(x_1, \cdots, x_r), y)$). On en conclut que z peut se mettre sous la forme $v \odot s_r$ (s_r étant identifié à l'application $((x_1, \cdots, x_r), y) \to (s_r(x_1, \cdots, x_r), y)$), v étant une fonction numérique sur $S^r \times U$. De plus, il est également clair que z est aussi fonction de définition de M en tout point de la forme $(p(a_1, \cdots, a_r), b)$, $p \in P$. Il résulte alors de la Proposition 2, § I que M se met sous la forme $s_r^*(D)$, D étant un diviseur sur $S^r \times U$, et que $\operatorname{Supp} M = s_r^{-1}(\operatorname{Supp} D)$; il résulte de cette dernière égalité que l'ensemble

[3] On reserve d'habitude le nom de système linéaire aux applications de la forme $\zeta \to \operatorname{div} \zeta + d_0$, où d_0 est un diviseur tel que $\operatorname{div} \zeta + d_0$ soit positif pour tout ζ. Il nous a semblé plus commode d'utiliser la terminologie donnée dans le texte.

$\{s_r(x_1, \cdots, x_r)\} \times U$ (où $(x_1, \cdots, x_r) \in C^r$) n'est jamais contenu dans $\operatorname{Supp} D$, de sorte que D définit une famille algébrique g de diviseurs de U paramétrée par S^r; il est clair que $m = g \circ s_r$.

On peut appliquer ce qui précède au cas où f est la famille de diviseurs de C paramétrée par C qui associe à tout $x \in C$ le diviseur, noté 1.x, auquel est associé le cycle 1.x; le diviseur de définition de cette famille est le diviseur de $C \times C$ auquel est associé le cycle constitué par la diagonale de $C \times C$ prise avec le coefficient 1. On voit donc qu'il existe une famille algébrique d_r de diviseurs de C paramétrée par S^r telle que

$$d_r(s_r(x_1, \cdots, x_r)) = x_1 + \cdots + x_r$$

pour tout $(x_1, \cdots, x_r) \in C^r$; nous dirons que d_r est la *famille canonique* de diviseurs de C paramétrée par S^r.

Si m est un diviseur quelconque de C, le cycle associé à m se met sous la form $\sum_{i=1}^{k} a_i x_i$, $x_1, \cdots, x_k$ étant des points mutuellement distincts de C; le nombre $\sum_{i=1}^{k} a_i$ s'appelle le *degré* de m; de plus, dans la démonstration qui va suivre, nous appellerons hauteur de m le nombre $\sum_{i=1}^{k} |a_i|$. Rappelons que si C est complète, tout diviseur principal sur C est de degré 0.

PROPOSITION 3. *Soit f une famille algébrique de diviseurs d'une courbe complète normale C paramétrée par une variété T; le degré de $f(t)$ est alors indépendant de t.*

Pour tout $r \geqq 0$, soit d_r la famille canonique de diviseurs de C paramétrée par S^r. Il est clair que, si m est un diviseur quelconque sur C, il y a des entiers $r \geqq 0$, $r' \geqq 0$ et des points $z \in S^r$, $z' \in S^{r'}$ tels que $m = d_r(z) - d_{r'}(z')$; on peut de plus supposer que $r + r'$ est égal à la hauteur de m, que nous noterons $h(m)$. Soient r, r' des entiers $\geqq 0$ quelconque; l'ensemble des points $(z, z', t) \in S^r \times S^{r'} \times T$ tels que $f(t) + d_{r'}(z') = d_r(z)$ est fermé (cf. Corollaire 4 au Théorème 1); l'image $H_{r,r'}$ de cet ensemble par la projection de $S^r \times S^{r'} \times T$ sur T est donc fermée ($S^r \times S^{r'}$ étant une variété complète). La variété T est la réunion des $H_{r,r'}$ pour tous les couples d'entiers $r \geqq 0$, $r' \geqq 0$; de plus, si les hauteurs des diviseurs $f(t)$, $t \in T$, restent bornées, au moins pour les points d'une partie dense de T, il y aura un nombre h tel que la réunion des $H_{r,r'}$ pour $r + r' \leqq h$ soit dense dans T, donc soit T tout entier; comme T est irréductible, il en résultera qu'il y a un couple (r, r') tel que $H_{r,r'} = T$, et $f(t)$ sera toujours de degré $r - r'$. Nous sommes donc ramenés à prouver que les hauteurs des diviseurs $f(t)$ restent bornées quand t parcourt les points d'une partie dense convenable de T.

Soit w une fonction numérique $\neq 0$ sur $T \times C$. Nous allons montrer qu'il y a une partie ouverte non vide T_1 de T et un entier h_1 tels que, pour $t \in T_1$, $j_t^*(\operatorname{div} w)$ soit défini et de hauteur $\leqq h_1$. Soient p et q les projections de $T \times C$ sur son premier et son second facteur; il y a alors des fonctions numériques $u_1, \cdots, u_n$ linéairement indépendantes sur C et des fonctions numériques θ_i, θ_i' sur T $(1 \leqq i \leqq n)$ telles que

$$w = \left(\sum_{i=1}^{n} (\theta_i \odot p)(u_i \odot q)\right)\left(\sum_{i=1}^{n} (\theta_i' \odot p)(u_i \odot q)\right)^{-1}.$$

On peut de plus supposer que $\theta_1 \neq 0$, $\theta_1' \neq 0$. Soit T_1 l'ensemble des points $t \in T$ tels que les fonctions θ_i, θ_i' soient définies en t et que $(\theta_1\theta_1')(t) \neq 0$; c'est une partie ouverte non vide de T, et, si $t \in T_1$, $j_t^*(\operatorname{div} w)$ est défini et égal au diviseur de la fonction $\left(\sum_{i=1}^{n} \theta_i(t) u_i\right)\left(\sum_{i=1}^{n} \theta_i'(t) u_i\right)^{-1}$. Il y a un diviseur $a \geqq 0$ sur C tel que $u_1, \cdots, u_n$ soient multiples de $-a$; si $c_1, \cdots, c_n$ sont des constantes non toutes nulles, $\operatorname{div} \sum_{i=1}^{n} c_i u_i$ se met sous la forme $a' - a$, où a' est un diviseur $\geqq 0$ de même degré que a, et est par suite de hauteur au plus égale au double du degré de a. Il en résulte que, si $t \in T_1$, $j_t^*(\operatorname{div} w)$ est de hauteur au plus égale au quadruple du degré de a.

Ceci étant, soit D le diviseur de définition de f. Il existe un recouvrement fini $(W_i)_{i \in I}$ de $T \times C$ par des ouverts $W_i \neq \emptyset$ tel que, pour chaque i, il existe une fonction numérique w_i sur $T \times C$ qui est fonction de définition de D en tout point de W_i. Pour chaque i, il existe une partie ouverte non vide T_i de T et un entier $h_i \geqq 0$ tels, pour tout $t \in T_i$, $j_t^*(\operatorname{div} w_i)$ soit défini et de hauteur $\leqq h_i$. L'intersection T' des T_i est une partie ouverte non vide de T. Soit t un point de cet ensemble; pour tout i, soit U_i l'ensemble des $x \in C$ tels que $(t, x) \in W_i$. Si x est un point de U_i, x intervient avec le même coefficient dans les cycles associés aux diviseurs $f(t)$ et $j_t^*(\operatorname{div} w_i)$, car $w_i \odot j_t$ est une fonction de définition de $f(t)$ en x. Il en résulte aussitôt que la hauteur de $f(t)$ est au plus égale à la somme des h_i pour tous les $i \in I$, ce qui établit la Proposition 3.

III. Criteres de rationalite (I). Nous allons indiquer une construction, essentiellement dûe à Cartier, qui permet d'associer à un diviseur sur un produit $T \times U$ et à un vecteur tangent à T un objet d'une nouvelle espèce, à savoir un diviseur additif sur U.

Soit U une variété; désignons par R le faisceau constant sur U dont les sections sur tout ouvert non vide U' sont les fonctions numériques sur U'.

Soit par ailleurs O le faisceau des anneaux locaux sur U; toute section du faisceau quotient R/O sur U s'appelle un *diviseur additif* sur U. Toute fonction numérique u sur U définit de manière évidente un diviseur additif, qu'on appelle le diviseur additif principal défini par u.

Ceci étant, soient T et U des variétés, D un diviseur, t un point de T tel que $j_t^*(D)$ soit défini (j_t étant l'application $x \to (t,x)$ de U dans $T \times U$) et L un vecteur tangent à T en t. Nous allons associer à D et à L un diviseur additif $< L, D >$ sur U. Soient p et q les projections de $T \times U$ sur T et sur U; pour chaque point $x \in U$, il y a un vecteur tangent bien determiné Λ_x à $T \times U$ en (t,x) dont les images par les applications dérivées de p et q en (t,x) sont L et 0 respectivement; nous dirons que $x \to \Lambda_x$ est le champ de vecteurs sur $\{t\} \times U$ défini par L. Soit w une fonction numérique sur $T \times U$ qui est définie en au moins un point de $\{t\} \times U$; pour tout $x \in U$ tel que w soit définie en (t,x), $< \Lambda_x, w >$ est un élément de K. Montrons que l'application $x \to < \Lambda_x, w >$ de l'ensemble ouvert des $x \in U$ tels que w soit définie en (t,x) se prolonge en une fonction numérique sur U, que nous désignerons par $< L, w >$. Soit x_0 un point de U tel que w soit définie en (t, x_0); w peut alors se mettre sous la forme $w'w''^{-1}$ où chacune des fonctions w', w'' est de la forme $\sum_{i=1}^{h} (\theta_i \odot p)(u_i \odot q)$, les θ_i étant des fonctions numériques sur T définies en t et les u_i des fonctions numériques sur U définies en x_0, et où, de plus, on a $w''(t, x_0) \neq 0$. Il existe un voisinage ouvert U_0 de x_0 dans U tel que chacune des fonctions u_i qui interviennent dans les expressions de w', w'' soit partout définie sur U_0 et que l'on ait $w''(t,x) \neq 0$ pour tout $x \in U_0$. Si x est un point de cet ensemble, on a

$$< \Lambda_x, w > = (w''(t,x))^{-2}(< \Lambda_x, w' > w''(t,x) - < \Lambda_x, w'' > w'(t,x)).$$

Par ailleurs, si $w_1 = \sum_{i=1}^{h} (\theta_i \odot p)(u_i \odot q)$, les θ_i étant définies et t et les u_i sur U_0, on a, pour $x \in U_0$, $< \Lambda_x, w_1 > = \sum_{i=1}^{h} < L, \theta_i > u_i(x)$. Il résulte de là que la restriction à U_0 de l'application $x \to < \Lambda_x, w >$ est une fonction numérique partout définie sur U_0. On en conclut que l'application $x \to < \Lambda_x, w >$ de l'ensemble U_1 des x tels que w soit définie en (t,x) est une fonction numérique partout définie sur U_1, ce qui démontre l'assertion faite plus haut.

On notera que, si h est un morphisme d'une variété T' dans la variété T et si L est l'image par la dérivée de h en un point $t' \in h^{-1}(t)$ d'un vecteur tangent L' à T' en t', on a $< L', w \odot h > = < L, w >$. En effet, considérant h comme définissant un morphisme de $T' \times U$ dans $T \times U$, le vecteur L'

définit un champ de vecteurs $(t', x) \to \Lambda'_x$ sur $T' \times U$ le long de $\{t'\} \times U$, et Λ_x n'est autre que l'image de Λ'_x par la dérivée de h en (t', x).

Ceci étant, soit (t, x) un point quelconque de $T \times U$, et soit w une fonction de définition de D en (t, x) ; w est alors définie en au moins un point de $\{t\} \times U$. La classe de la fonction $(w \odot j_t)^{-1} < L, w >$ modulo l'anneau local $\mathfrak{o}(x)$ de x ne dépend pas du choix de w. En effet, si w_1 est une autre fonction de définition de D en (t, x), $w^{-1}w_1 = z$ est une fonction définie en (t, x) et y prenant une valeur $\neq 0$, de sorte que $(z \odot j_t)^{-1} < L, z >$ est dans $\mathfrak{o}(x)$; or on a

$$(w_1 \odot j_t)^{-1} < L, w_1 > = (w \odot j_t)^{-1} < L, w > + (z \odot j_t)^{-1} < L, z >,$$

ce qui établit notre assertion. Désignons par δ_x la classe de $(w \odot j_t)^{-1} < L, w >$ modulo $\mathfrak{o}(x)$; il est clair que, pour tous les points x' d'un voisinage convenable de x, $\delta_{x'}$ est aussi la classe de $(w \odot j_t)^{-1} < L, w >$ modulo $\mathfrak{o}(x')$. Ceci montre que l'application $x \to \delta_x$ est un diviseur additif; nous le noterons $< L, D >$. Il est clair que l'on a

$$< L, D + D' > = < L, D > + < L, D' >$$

si D et D' sont des diviseurs sur $T \times U$ tels que $j_t^*(D)$ et $j_t^*(D')$ soient définis. Si w est une fonction numériqe sur $T \times U$ définie en au moins un point de $\{t\} \times U$, $< L, \operatorname{div} w >$ est le diviseur additif associé à la fonction numérique $(w \odot j_t)^{-1} < L, w >$ sur U.

LEMME 1. *Soient T, T', U variétés, h un morphisme de T' dans T, t' un point de T', t le point $h(t')$, L' un vecteur tangent à T' en t', L l'image de L' par l'application dérivée de h, D un diviseur sur $T \times U$ tel que $h^*(D)$ et $j_t^*(D)$ soient définis (j_t étant l'application $x \to (t, x)$ de U dans $T \times U$). On a alors $< L', h^*(D) > = < L, D >$.*

Si $x \in U$ et si w est fonction de définition de D en (t, x), $w \odot h$ est fonction de définition de $h^*(D)$ en (t', x) ; le Lemme 1 résulte alors de ce qui a été dit plus haut.

Si Δ est un diviseur additif sur U, la valeur de Δ en un point x est une classe modulo l'anneau local de x; tout représentant de cette classe s'appelle une fonction de définition de Δ en x.

LEMME 2. *Soient T et U des variétés, t un point de T, L un vecteur tangent à T en t, D un diviseur $\geqq 0$ sur $T \times U$ tel que $j_t^*(D)$ soit défini. Soient u une fonction de définition de $j_t^*(D)$ en un point $x \in U$ et s une fonction de définition du diviseur additif $< L, D >$ en x; la fonction su est alors définie en x.*

Soit w une fonction de définition de D en (t, x) ; on peut supposer que $u = w \odot j_t$ et que $s = (w \odot j_t)^{-1} < L, w >$; le lemme résulte alors de ce que, w étant définie en (t, x), $< L, w >$ est définie en x.

Si f est une famille algébrique de diviseurs d'une variété U paramétrée par une variété T, si $t \in T$ et si L est un vecteur tangent à T en t, nous poserons $< L, f > = < L, D >$, D étant le diviseur de définition de f. Si on a $< L, f > \neq 0$ pour tout vecteur tangent $L \neq 0$ à T en t, on dit que f est *infinitésimalement injective en* t. Si f est infinitésimalement injective en tout point de T, on dit que cette famille est *infinitésimalement injective.*

PROPOSITION 2. *Soit f une famille injective de diviseurs d'une variété U paramétrée par une variété complète T. Soit f' une famille de diviseurs de U paramétrée par une variété normale T' ; supposons que, pour tout $t' \in T'$ il existe un point $t \in T$ tel que $f(t) = f'(t')$ et qu'il existe un point $(t_0, t'_0) \in T \times T'$ tel que $f(t_0) = f'(t'_0)$ et que f soit infinitésimalement injective en t_0. Il existe alors un morphisme h de T' dans T tel que $f' = f \circ h$.*

L'ensemble E des points $(t, t') \in T \times T'$ tels que $f(t) = f'(t')$ est fermé (Corollaire 4 au Théorème 1, § I), et il résulte des hypothèses que la projection $T \times T' \to T'$ induit une bijection p de E sur T'. Montrons que E est irréductible. Il existe au moins une composante irréductible E_1 de E telle que $p(E_1)$ soit dense dans T'. Or, T étant complète, la projection $T \times T' \to T'$ est une application propre ; comme E_1 est fermé, on a $p(E_1) = T'$, d'où $E_1 = E$, puisque p est bijectif, ce que établit notre assertion. L'application p est donc un morphisme bijectif propre de E sur T' ; si nous montrons qu'il est birationnel, il résultera du théorème principal de Zariski et du fait que T' est normale que p est un isomorphisme de E sur T'. En composant l'isomorphisme p^{-1} de T' sur E avec la restriction à E de la projection $T \times T' \to T$, on obtiendra un morphisme h possédant la propriété requise.

Or, le morphisme p, qui est bijectif, est radiciel ; pour montrer qu'il est birationnel, il suffira de montrer qu'il est séparable, donc qu'il y a un point de E en lequel p n'est pas ramifié (i. e. tel que l'application dérivée de p en ce point soit injective). Nous allons voir que le point (t_0, t'_0) possède cette propriété. Soit L'' un vecteur tangent $\neq 0$ à E en (t_0, t'_0) ; soient L et L' les images de L'' par les applications dérivées des restrictions q et p à E des projections de $T \times T'$ sur T et sur T' ; comme L'' s'identife à un vecteur tangent à $T \times T'$, L et L' ne sont pas tous deux nuls ; nous voulons montrer que $L \neq 0$. Il est clair que l'on a $f \circ q = f' \circ p$; soit f'' leur valeur comme. Il résulte du Lemme 1 que $< L'', f'' > = < L, f > = < L', f' >$; si on avait $L' = 0$, on aurait $L \neq 0$, d'où $< L, f > \neq 0$ puisque f est infinitésimalement injective en t_0, d'où contradiction. La Proposition 2 est donc établie.

Nous allons maintenant donner des exemples de familles infinitésimalement injectives. Soit U une variété *semi-complète,* et soit V un sous-espace vectoriel de dimension finie > 0 de l'espace des fonctions numériques sur U; V définit alors un système linéaire f de diviseurs de U paramétré par $\mathfrak{P}(V)$. Nous allons voir que f est infinitésimalement injectif. Désignons par V_0 l'ensemble des éléments $\neq 0$ de V et par φ l'application canonique de V_0 sur $\mathfrak{P}(V)$; si $u_0 \in V_0$, l'application dérivée de φ en u_0 est une surjection de l'espace tangent à V_0 en u_0 sur l'espace tangent à $\mathfrak{P}(V)$ en $\varphi(u_0)$. Tenant compte du Lemme 1, on voit qu'il suffira de montrer que tout vecteur tangent L à V_0 en u_0 tel que $< L, f\circ\varphi > = 0$ appartient au noyau de la dérivée de φ. Or le diviseur de définition de $f\circ\varphi$ est $\operatorname{div} w$, où w est la fonction numérique sur $V_0 \times U$ telle que $w(u, x) = u(x)$ si $u \in V_0$ et si u est définie en x. Soit $(u_0, u_1, \cdots, u_m)$ une base de V contenant u_0; soient $\lambda_0, \cdots, \lambda_m$ les coordonnées relativement à cette base; si donc x est un point en lequel $u_0, \cdots, u_m$ sont définies, on a $w(u, x) = \sum_{i=0}^{m} \lambda_i(u) u_i(x)$, d'où il résulte tout de suite que $< L, w >$ est la fonction $\sum_{i=0}^{m} < L, \lambda_i > u_i$, et que $< L, \operatorname{div} w >$ est le diviseur additif représenté par la fonction $\sum_{i=0}^{m} < L, \lambda_i > u_0^{-1}u_i$. Si ce diviseur additif est nul, la fonction $\sum_{i=0}^{m} < L, \lambda_i > u_0^{-1}u_i$ est partout définie, donc constante puisque U est semi-complète; comme $u_0, \cdots, u_m$ sont linéairement indépendantes, ceci n'est possible que si $< L, \lambda_i > = 0$ pour $0 \leqq i \leqq m$; or ceci est précisément la condition pour que l'image de L par la dérivée de φ soit nulle.

Soit maintenant C une courbe normale, et soit r un entier > 0; soit d_r la famille canonique de diviseurs de C paramétrée par la puissance symétrique r-ième S^r de C, et soit s_r l'application canonique de C^r sur S^r. Nous allons montrer que d_r est infinitésimalement injective en tout point de S^r de la forme $s_r(a_1, \cdots, a_r)$, $a_1, \cdots, a_r$ étant des points *tous distincts* de C [en fait, on peut montrer que d_r est infinitésimalement injective en tout point de S^r, mais le raisonnement est plus compliqué]. Le morphisme s_r n'est pas ramifié au point $(a_1, \cdots, a_r)$, qui est simple sur C^r; sa dérivée en ce point est donc un isomorphisme de l'espace tangent à C^r en $(a_1, \cdots, a_r)$ sur l'espace tangent à S^r au point $s_r(a_1, \cdots, a_r)$. Il nous suffira donc de montrer que l'on a $< L, d_r \circ s_r > \neq 0$ si L est un vecteur tangent $\neq 0$ à C^r en $(a_1, \cdots, a_r)$. Soient $q_1, \cdots, q_r, q_{r+1}$ les projections de $C^{r+1} = C^r \times C$ sur les divers facteurs. Soient $L_1, \cdots, L_r$ les images de L par les dérivées des projections de C^r sur ses divers facteurs; L_i est donc un vecteur tangent à C en a_i, et il existe au moins un i, soit k, tel que $L_i \neq 0$. Soit u une variable uniformisante en a_k sur C. L'application $d_r \circ s_r$ est l'application $(x_1, \cdots, x_r) \to \sum_{i=1}^{r} x_i$; tenant

compte de ce que $a_1, \cdots, a_r$ sont mutuellement distincts, on voit tout de suite que la fonction $w = u \odot q_k - u \odot q_{r+1}$ est fonction de définition en $((a_1, \cdots, a_r), a_k)$ du diviseur de définition de $d_r \circ s_r$. Il est clair que $< L, w >$ est la fonction constante $< L_k, u >$ sur C, de sorte que la fonction $(u(a_k) - u)^{-1} < L_k, u >$ est fonction de définition de $< L, d_r \circ s_r >$ en a_k. Or, on a $< L_k, u > \neq 0$ puisque u est variable uniformisante en a_k et $L_k \neq 0$; comme a_k est un zéro de $u(a_k) - u$, on voit que $< L, d_r \circ s_r > \neq 0$, ce qui établir notre assertion.

IV. Familles algebriques de classes de diviseurs. Pour toute variété X, nous désignerons par $\mathfrak{D}(X)$ le groupe des diviseurs de X, par $\mathfrak{P}(X)$ le groupe des diviseurs principaux de X et par $\mathfrak{G}(X) = \mathfrak{D}(X)/\mathfrak{P}(X)$ le groupe des classes de diviseurs de X.

PROPOSITION 1. *Soit $\mathfrak{k}$ une classe de diviseurs sur une variété X; si x_0 est un point de X, il existe dans $\mathfrak{k}$ un diviseur dont le support ne contient pas x_0.*

Soient en effet d un diviseur quelconque de la classe $\mathfrak{k}$ et u une fonction de définition de d en x_0; $d' = d - \operatorname{div} u$ possède alors la propriété requise.

Soit f un morphisme d'une variété V dans une variété U. Si $\mathfrak{k} \in \mathfrak{G}(U)$, il existe toujours un diviseur $d \in \mathfrak{k}$ tel que $f^*(d)$ soit défini; il suffit en effet de choisir un point $y_0 \in V$ et un diviseur $d \in \mathfrak{k}$ tel que $f(y_0) \notin \operatorname{Supp} d$. De plus, les diviseurs $f^*(d)$, pour tous les diviseurs $d \in \mathfrak{k}$ tels que $f^*(d)$ soit défini, appartiennent tous à une même classe. Pour le voir, il suffit de montrer que, si d est un diviseur principal tel que $f^*(d)$ soit défini, $f^*(d)$ est principal. Or, si $d = \operatorname{div} u$, il y a au moins un point $y \in V$ tel que u et u^{-1} soient définies en $f(y)$, de sorte que $u \odot f$ est définie et $\neq 0$; il résulte alors tout de suite des définitions que $f^*(d) = \operatorname{div} u \odot f$. Nous désignerons par $f^*(\mathfrak{k})$ la classe de diviseurs de V qui contient les $f^*(d)$ pour les $d \in \mathfrak{k}$ tels que $f^*(d)$ soit défini. Il est clair que l'application f^* ainsi définie est un homomorphisme de $\mathfrak{G}(U)$ dans $\mathfrak{G}(V)$. Soit maintenant g un morphisme d'une variété W dans la variété V; on a alors $g^*(f^*(\mathfrak{k})) = (f \circ g)^*(\mathfrak{k})$ pour tout $\mathfrak{k} \in \mathfrak{G}(U)$. En effet, soit d un représentant de $\mathfrak{k}$ tel que $(f \circ g)^*(d)$ soit défini; alors $f^*(d)$ et $g^*(f^*(d))$ sont définis, et $g^*(f^*(d)) = (f \circ g)^*(d)$, ce qui démontre notre assertion. Il résulte de là que l'application $U \to \mathfrak{G}(U)$ définit un foncteur contravariant sur la catégories des variétés à valeurs dans celle des groupes abéliens.

Soient en particulier T et U des variétés, et $\mathfrak{k}$ un élément de $\mathfrak{G}(T \times U)$. Pour tout $t \in T$, soit j_t l'application $x \to (t, x)$ de U dans $T \times U$; l'application

$f: t \to j_t^*(\mathfrak{k})$ est alors une application de T dans $\mathfrak{G}(U)$. On appelle *famille algébrique des classes de diviseurs de U paramétrée par T* (ou application algébrique de T dans $\mathfrak{G}(U)$) toute application de T dans $\mathfrak{G}(U)$ qui peut se définir de la manière qu'on vient d'indiquer à partir d'une classe de diviseurs sur $T \times U$.

Pour tout diviseur d sur une variété, nous désignerons par Cl. d la classe de diviseurs de d. Si $\bar{f}$ est une famille algébrique de diviseurs d'une variété U paramétrée par une variété T, l'application $t \to \mathrm{Cl}.\bar{f}(t)$ est évidemment une famille algébrique de classes de deviseurs.

PROPOSITION 2. *Soit f une famille algébrique de classes de diviseurs d'une variété U paramétrée par une variété T. Si $t_0 \in T$, il existe un voisinage ouvert T_0 de t_0 et une famille algébrique $\bar{f}$ de diviseurs de U paramétrée par T_0 tels que l'on ait $f(t) = \mathrm{Cl}.\bar{f}(t)$ pour tout $t \in T_0$.*

Soit x_0 un point de U. La famille f est définie par une classe $\mathfrak{k} \in \mathfrak{G}(T \times U)$, et $\mathfrak{k}$ contient un diviseur D dont le support ne passe pas par (t_0, x_0). Il existe un voisinage ouvert T_0 de t_0 tel que $T_0 \times \{x_0\}$ ne rencontre pas Supp D; soit D_0 le diviseur induit par D sur $T_0 \times U$. Il est clair que, si $t \in T_0$, $\{t\} \times U$ n'est pas contenu dans Supp D_0; D_0 définit donc une famille algébrique $\bar{f}$ de diviseurs de U paramétrée par T_0, qui possède évidemment la propriété requise.

Observons maintenant que, si une classe $\mathfrak{k} \in \mathfrak{G}(T \times U)$ définit une famille f de classes de diviseurs de U, il n'est pas nécessaire que l'on ait $\mathfrak{k} = 0$ pour que l'on ait $f = 0$. Soit en effet p la projection de $T \times U$ sur son premier facteur; supposons que $\mathfrak{k} \in p^*(\mathfrak{G}(T))$; on va montrer que l'on a alors $f = 0$. Soit $\mathfrak{k} = p^*(\mathfrak{k}_1)$, $\mathfrak{k}_1 \in \mathfrak{G}(T)$. Soit $t \in T$; il y a dans $\mathfrak{k}_1$ un diviseur d_1 dont le support ne contient pas t; $d_1 \times U$ est alors un représentant de $\mathfrak{k}$; j_t^* étant défini comme plus haut, $j_t^*(d_1 \times U)$ est défini et nul puisque $t \notin \mathrm{Supp}\, d_1$; il en résulte que $f(t) = 0$ quel que soit t. Ceci conduit à introduire le groupe

$$\mathfrak{M}(T;U) = \mathfrak{G}(T \times U)/p^*(\mathfrak{G}(T));$$

si m est un élément de ce groupe, toutes les classes $\mathfrak{k} \in m$ définissent la même famille f de classes de diviseurs de U; on dit que f est la famille définie par m. Nous verrons tout à l'heure que, sous certaines hypothèses, la condition $f = 0$ entraîne $m = 0$.

Si h est un morphisme d'une variété T' dans une variété T, et si on désigne par p' la projection de $T' \times U$ sur $T \times U$, il est clair que l'homomorphisme h^* de $\mathfrak{G}(T \times U)$ dans $\mathfrak{G}(T' \times U)$ applique $p^*(\mathfrak{G}(T))$ dans $p'^*(\mathfrak{G}(T'))$, et définit par suite un homomorphisme de $\mathfrak{M}(T, U)$ dans $\mathfrak{M}(T', U)$.

Désignons encore par q la projection de $T \times U$ sur son deuxième facteur U. Il nous sera commode d'introduire aussi le groupe

$$\mathfrak{N}(T, U) = \mathfrak{G}(T \times U) / (p^*(\mathfrak{G}(T)) + q^*(\mathfrak{G}(U)))\,;$$

ici encore, à tout morphisme d'une variété T' dans la variété T est attaché un homomorphisme de $\mathfrak{N}(T, U)$ dans $\mathfrak{N}(T', U)$, de sorte que, pour U fixe, les groupes $\mathfrak{N}(T, U)$ pour toutes les variétés T définissent un facteur contravariant sur la catégorie des variétés. On notera que l'isomorphisme canonique de $U \times T$ sur $T \times U$ définit un *isomorphisme canonique*

$$\mathfrak{N}(T, U) \cong \mathfrak{N}(U, T).$$

On notera que, si une classe $\mathfrak{k}$ de diviseurs de $T \times U$ appartient au groupe $q^*(\mathfrak{G}(U))$, l'application algébrique f de T dans $\mathfrak{G}(U)$ qu'elle définit est constante; en effet, $\mathfrak{k}$ contient alors un diviseur de la forme $T \times d$, où d est un diviseur de U; si t est un point quelconque de T, $f(t)$ est la classe du diviseur d.

THÉORÈME 2. *Soit f une famille algébrique de classes de diviseurs d'une variété semi-complète U paramétrée par une variété T. L'ensemble des points $t \in T$ tels que $f(t) = 0$ est alors fermé.*

Nous établirons d'abord le lemme suivant:

LEMME 1. *Soient T et U des variétés, $\mathfrak{k}$ une classe de diviseurs de $T \times U$, t_0 un point de T. Il existe alors un diviseur D de la classe $\mathfrak{k}$, un voisinage ouvert T_0 de t_0 dans T et un faisceau A d'idéaux fractionnaires sur U qui possèdent les propriétés suivantes: si $t \in T_0$, $j_t^*(D) = \delta(t)$ est défini (j_t étant l'application $x \to (t, x)$ de U dans $T \times U$), et le faisceau d'idéaux $A^{\delta(t)}$ est un sous-faisceau de A.*

Soit d'abord D_1 un représentant quelconque de la classe $\mathfrak{k}$ tel que $j_{t_0}^*(D_1)$ soit défini. Il existe un voisinage affine T_1 de t_0 tel que $j_t^*(D_1)$ soit défini pour tout $t \in T_1$. Soit U_0 un morceau affine quelconque de U. Les sections sur $T_1 \times U_0$ du faisceau d'idéaux fractionnaires sur $T \times U$ associé à D_1 forment un idéal fractionnaire $\mathfrak{D}$ pour l'algèbre affine $\mathfrak{O}$ de $T_1 \times U_0$. Soit $\mathfrak{X}$ l'idéal de $\mathfrak{O}$ composé des éléments $w \in \mathfrak{O}$ tels que $w\mathfrak{D} \subset \mathfrak{O}$. Si $(t, x) \in T_1 \times U_0$, l'idéal engendré par $\mathfrak{X}$ dans l'anneau local $\mathfrak{o}(t, x)$ de (t, x) n'est autre que l'ensemble des éléments $w \in \mathfrak{o}(t, x)$ tels que $w(\mathfrak{D}\mathfrak{o}(t, x)) \subset \mathfrak{o}(t, x)$. Puisque $j_{t_0}^*(D_1)$ est défini, il y a un point $x_0 \in U_0$ tel que $\mathfrak{D}\mathfrak{o}(t_0, x_0) = \mathfrak{o}(t_0, x_0)$; il y a donc un élément $w \in \mathfrak{X}$ tel que $w(t_0, x_0) \neq 0$. Posons $D = D_1 + \operatorname{div} w$; D est encore un représentant de la classe $\mathfrak{k}$; il est clair que D est positif en

tout point de $T_1 \times U_0$; de plus, $j_{t_0}(D)$ est défini. Il y a donc un voisinage ouvert T_0 de t_0 tel que $j_t{}^*(D)$ soit défini toutes les fois que $t \in T_0$. Soit D' le diviseur induit par D sur $T_0 \times U$; si nous posons $E = T_0 \times (U - U_0)$, D' est positif en tout point de $T_0 \times (U - E)$. Si donc B est le faisceau d'idéaux qui définit l'ensemble fermé E, il y a un entier $k \geqq 0$ tel que $B^k A^{D'}$ soit un faisceau d'idéaux entiers, $A^{D'}$ désignant le faisceou d'idéaux sur $T_0 \times U$ associé à D' (cf. § I). Désignons par C le faisceou d'idéaux sur U qui définit l'ensemble $U - U_0$, et par A le transporteur de C^k dans le faisceau $\mathfrak{O}$ des anneaux locaux de U. Nous allons montrer que, si $t \in T_0$, $\delta(t) = j_t{}^*(D)$, le faisceau d'idéaux associé à $\delta(t)$ est un sous-faisceau de A. Soient x un point de U, w une fonction de définition de D en (t, x) et $u = w \odot j_t$; u est donc fonction de définition de $\delta(t)$ en x. Pour montrer que l'idéal ponctuel en x du faisceau associé à $\delta(t)$ est contenu dans A_x, il suffit de montrer que, si $v_1, \cdots, v_k$ sont des fonctions de l'idéal ponctuel C_x de C en x, $v_1 \cdots v_k u$ appartient à l'anneau local de x. Soit q la projection $T \times U \rightarrow U$; chacune des fonctions v_i est nulle sur l'intersection avec $U - U_0$ d'un voisinage convenable de x, de sorte que les fonctions $v_i \odot q$ sont nulles sur l'intersection de E avec un voisinage convenable de (t, x), ce qui montre qu'elles appartiennent à l'idéal ponctuel de B en (t, x); la fonction $(v_1 \cdots v_k \odot q) w$ est donc définie en (t, x). Or on a $v_i = (v_i \odot q) \odot j_t$, d'où

$$v_1 \cdots v_k u = ((v_1 \cdots v_k \odot q) w) \odot j_t,$$

ce qui établit notre assertion. Le Lemme 1 est donc établi.

Ceci établi, nous pouvons démontrer le Théorème 2. L'application f est définie au moyen d'une classe $\mathfrak{f}$ de diviseurs de $T \times U$; soit t_0 un point de T adhérent à l'ensemble H des points t tels que $f(t) = 0$; nous choisirons T_0, D, A comme dans le Lemme 2. Puisque U est semi-complète, les sections de A forment un espace vectoriel V de dimension finie. Si $t \in H \cap T_0$, $\delta(t)$, qui est un représentant de la classe $f(t)$, est un diviseur principal; c'est donc le diviseur d'une fonction qui appartient à V. Ceci montre que $V \neq \{0\}$; soit $\zeta \rightarrow \operatorname{div} \zeta$ le système linéaire paramétrée par l'espace projectif $\mathfrak{P}(V)$ associé à V. Alors $H \cap T_0$ est l'image par la projection de $T_0 \times \mathfrak{P}(V)$ sur T_0 de l'ensemble L des couples (t, ζ) tels que $f(t) = \operatorname{div} \zeta$ (car, pour tout $\zeta \in \mathfrak{P}(V)$, $\operatorname{div} \zeta$ est principal). Or, L est fermé (Corollaire 4 au Théorème 1, § II); $H \cap T_0$ est donc relativement fermé dans T_0, d'où $t_0 \in H$, ce qui démontre le Théorème 2.

THÉORÈME 3. *Soient T une variété normale, U une variété semi-complète et m un élément de $\mathfrak{M}(T, U)$; la famille algébrique de classes de diviseurs définie par m ne peut être nulle que si $m = 0$.*

Reprenons les notations de la démonstration du Théorème 2, t_0 étant ici un point quelconque de T. On a vu au § III que le système linéaire $\zeta \to \operatorname{div} \zeta$ est une famille injective et infinitésimalement injective. Il existe donc un morphisme g de T_0 dans $\mathfrak{P}(V)$ tel que l'on ait $\delta(t) = \operatorname{div} g(t)$ pour $t \in T_0$. Soit φ l'application canonique de l'ensemble V_0 des éléments $\neq 0$ de V sur $\mathfrak{P}(V)$; il existe un morphisme ψ d'un voisinage P de $g(t_0)$ dans V_0 tel que $\varphi \circ \psi$ soit l'application identique de P. Soit T_0' un voisinage ouvert de t_0 tel que $g(T_0') \subset P$, et soit g' l'application $t \to \psi(g(t))$ de T_0' dans V_0 ; on a donc $\delta(t) = \operatorname{div} g'(t)$ si $t \in T_0'$ (rappelons que $g'(t)$, en tant qu'élément de V_0, est une fonction numérique sur U). Si U_1 est l'ensemble des points de U en lesquels sont définies toutes les fonctions de V, il est clair que $(t,x) \to (g'(t))(x)$ est une fonction numérique sur $T_0' \times U_1$ qui se prolonge en une fonction numérique w sur $T \times U$; si $t \in T_0'$ on a

$$j_t^*(\operatorname{div} w) = \operatorname{div} g'(t) = \delta(t) = j_t^*(D),$$

j_t désignant l'application $x \to (t,x)$ de U dans $T \times U$. On en conclut que le diviseur induit par $D - \operatorname{div} w$ sur $T_0' \times U$ définit la famille nulle de diviseurs de U paramétrée par T_0', donc est nul. Or $D' = D - \operatorname{div} w$ est un représentant de la classe $\mathfrak{k}$ de diviseurs de $T \times U$ qui contient D. Il résulte de là que $\operatorname{Supp} D' \subset (T - T_0') \times U$. Le Théorème 3 résultera donc du

LEMME 2. *Soient T une variété normale, U une variété et D un diviseur sur $T \times U$. S'il existe une partie fermée $E \neq T$ de T telle que* $\operatorname{Supp} D$ $\subset E \times U$, *D est de la forme $d \times U$, d étant un diviseur de T.*

Nous considérerons d'abord le cas où U est aussi supposée normale. Pour tout point $x \in U$, nous désignerons par k_x l'application $t \to (t,x)$ de T dans $T \times U$. Nous allons montrer que, si $D \neq 0$, on a $k_x^*(D) \neq 0$ pour tout $x \in U$. Puisque $T \times U$ est normale, toute composante irréductible Σ de $\operatorname{Supp} D$ est une hypersurface de $T \times U$; comme $\Sigma \subset E \times U$, il est clair que Σ est de la forme $S \times U$, S étant une hypersurface de T. Soient $S_1, \cdots, S_h$ les hypersurfaces de T telles que $S_i \times U \subset \operatorname{Supp} D$ (avec $S_i \neq S_j$ si $i \neq j$) ; soit t un point de S_1 qui n'appartient à aucun S_i d'indice $i > 1$. Soit w une fonction de définition de D en (t,x) ; $S_1 \times U$ est donc la seule hypersurface de $T \times U$ passant part (t,x) qui soit ou bien variété de pôles ou bien variété de zéros de w. La variété $T \times U$ étant normale, il en résulte que, si (t,x) est pôle de w, w^{-1} est définie en (t,x) et y prend la valeur 0, tandisque, si (t,x) est un zéro de w, w est définie en ce point et y prend la valeur 0. On en conclut que, dans le premier cas, t est un pôle de $w \odot k_x$, tandisque, dans le second cas, c'en est un zéro; dans les deux cas, t appartient à $k_x^*(D)$. Ceci

étant, soit x_0 un point quelconque de U; posons $d = k_{x_0}^*(D)$, $D' = D - d \times U$; il est clair que $\operatorname{Supp} D' \subset (E \cup \operatorname{Supp} d) \times U$; de plus, on a $k_{x_0}^*(D') = 0$; en vertu de ce qu'on vient d'établir, il en résule que $D' = 0$. Pour passer au cas général, désignons par U_0 l'ensemble des points de U en lesquels U est normale; c'est une sous-variété ouverte et normale de U. Il en résulte que le diviseur D_0 induit par D sur $T \times U_0$ se met sous la forme $d \times U_0$, d étant un diviseur sur T; on a donc $k_x^*(D - d \times U) = 0$ pour tout $x \in U_0$. Or, $x \to k_x^*(D - d \times U)$ est une famille algébrique de diviseurs de T paramétrée par U; comme sa restriction à la partie dense U_0 de U est nulle, son diviseur de définition $D - d \times U$ est nul, ce qui démontre le Lemme 2.

Soient T une variété normale et U une variété; désignons par p la projection $T \times U \to T$. Si T' et T'' sont des parties ouvertes non vides de T avec $T' \subset T''$, l'injection canonique $T' \to T''$ définit des homomorphismes $\mathfrak{D}(T'' \times U) \to \mathfrak{D}(T' \times U)$, $\mathfrak{P}(T'' \times U) \to \mathfrak{P}(T' \times U)$, $\mathfrak{D}(T'') \to \mathfrak{D}(T')$, $p^*(\mathfrak{D}(T'')) \to p^*(\mathfrak{D}(T'))$. Les applications

$$T' \to \mathfrak{D}(T' \times U),\ T' \to \mathfrak{P}(T' \times U),\ T' \to p^*(\mathfrak{D}(T'))$$

sont donc munies de structures de pré-faisceaux de groupes commutatifs sur T. Il est évident que l'application $T' \to \mathfrak{D}(T' \times U)$ est un faisceau, que nous désignerons par $\mathfrak{D}_U$. Montrons que l'application $T' \to \mathfrak{P}(T' \times U) + p^*(\mathfrak{D}(T'))$ est un sous-faisceau de $\mathfrak{D}_U$. Sachant déjà que $\mathfrak{D}_U$ est un faisceau, il suffit de vérifier ce qui suit: soit D' un diviseur sur $T' \times U$; supposons qu'il existe un recouvrement $(T'_i)_{i \in I}$ de T' par des ouverts non vides T'_i tels que, pour tout i, le diviseur D'_i induit par D' sur $T'_i \times U$ soit de la forme $\operatorname{div} w_i + d' \times U$, où w_i est une fonction numérique sur $T'_i \times U$ et d'_i un diviseur sur T'_i; alors D est de la forme $\operatorname{div} w' + d'$, w' étant une fonction numérique sur $T' \times U$ et d' un diviseur sur T'. Chacune des fonctions w_i se prolonge en une fonction numérique sur $T' \times U$, que nous désigerons encore par w_i. Soit i_0 un élément de I; remplaçant D' par $D' - \operatorname{div} w_{i_0}$, nous nous ramenons au cas où $\operatorname{div} w_{i_0} = 0$. Pour chaque i, le diviseur induit par D'_i sur $(T'_i \cap T'_{i_0}) \times U$ est alors identique à celui qui est induit par $d'_{i_0} \times U$ sur ce même ensemble; il en résulte que, si E_i est la réunion de $\operatorname{Supp} d'_{i_0}$ et de $T' - (T'_i \cap T'_{i_0})$, on a $\operatorname{Supp} D'_i \subset E_i \times U$, d'où il résulte, en vertu du Lemme 2, que D'_i est de la forme $d''_i \times U$, d''_i étant un diviseur sur T'_i. Il est clair que, pour toute sous-variété ouverte T'' de T, l'application $d'' \to d'' \times U$ de $\mathfrak{D}(T'')$ dans $\mathfrak{D}(T'' \times U)$ est injective; on en conclut que, si $i, j \in I$, d''_i et d''_j induisent le même diviseur sur $T'_i \cap T'_j$. Il existe donc un diviseur d' sur T' tel que, pour tout i, d'_i soit le diviseur induit par d' sur T'_i. Le diviseur $D' - d' \times U$ induit alors 0 sur chacun des $T'_i \times U$ et est par suite nul, ce qui démontre notre assertion.

Si T' est une partie ouverte non vide de T, on a

$$\mathfrak{M}(T',U)=\mathfrak{D}(T'\times U)/(\mathfrak{P}(T'\times U)+p^*(\mathfrak{D}(T'))).$$

L'application $T'\to\mathfrak{M}(T',U)$ est évidemment munie d'une structure de pré-faisceau; ce pré-faisceau est le quotient *dans la catégorie des pré-faisceaux* du faisceau $\mathfrak{D}_U$ par le faisceau $\mathfrak{Q}_U$: $T'\to\mathfrak{P}(T'\times U)+p^*(\mathfrak{D}(T'))$. Il n'en résulte pas que $T'\to\mathfrak{M}(T',U)$ soit un faisceau. Cependant, observons que, si T est une variété non singulière, $\mathfrak{Q}_U$ est un faisceau *flasque*. En effet, soient alors T' une partie ouverte non vide de T, w' une fonction numérique sur $T'\times U$ et d' un diviseur sur T'; w' se prolonge alors en une fonction numérique w sur $T\times U$, et $\operatorname{div} w'$ est le diviseur induit par $\operatorname{div} w$ sur $T'\times U$; par ailleurs, T étant non singulière, d' est induit sur T' par un diviseur d sur T (Corollaire à la Proposition 1, § I); il en résulte que $\operatorname{div} w'+d'\times U$ est le diviseur induit par $\operatorname{div} w+d\times U$ sur $T'\times U$, ce qui montre bien que $\mathfrak{Q}_U$ est flasque. Or le pré-faisceau quotient d'un faisceau par un sous-faisceau flasque est un faisceau ([4], Théorème 3. 1. 2, p. 148); on voit donc que, si T est non singulière, $T'\to\mathfrak{M}(T',U)$ est un faisceau de groupes commutatifs. Si on suppose U semi-complète, il y a correspondance biunivoque entre éléments de $\mathfrak{M}(T,U)$ et familles algébriques de classes de diviseurs paramétrées par T; on obtient donc le résultat suivant:

PROPOSITION 3. *Soient T une variété non singulière et U une variété semi-complète; soit f une application de T dans $\mathfrak{G}(U)$. Supposons que tout point $t\in T$ admette un voisinage ouvert T' tel que la restriction de f à T' soit une famille algébrique paramétrée par T'; f est alors une famille algébrique paramétrée par T.*[4]

COROLLAIRE. *Les notations étant celles de la Proposition 3, supposons que tout point de T admette un voisinage ouvert T' qui possède la propriété suivante: il existe une famille algébrique f' de diviseurs de U paramétrée par T' telle que $f(t)=\mathrm{Cl}.f'(t)$ pour tout $t\in T'$. Alors f est une famille algébrique paramétrée par T.*

Soient T et U des variétés, p et q les projections de $T\times U$ sur T et sur U. Alors $q^*(\mathfrak{G}(U))$ est un sous-groupe de $\mathfrak{G}(T\times U)$; désignons par $\mathfrak{M}_0(T,U)$ son image canonique dans $\mathfrak{M}(T,U)=\mathfrak{G}(T\times U)/p^*(\mathfrak{G}(T))$; on a donc $\mathfrak{N}(T,U)=\mathfrak{M}(T,U)/\mathfrak{M}_0(T,U)$; de plus, on a une application surjective $\mathfrak{G}(U)\to\mathfrak{M}_0(T,U)$; montrons que cette application est même un isomorphisme. Il suffit de montrer que, si $\mathfrak{c}$ est une classe de diviseurs de U telle

[4] Serre m'a communiqué une démonstration du fait que l'hypothèse " T non singulière " est superflue dans l'énoncé de la Proposition 3.

que $q^*(\mathfrak{c})$ appartienne à $p^*(\mathfrak{G}(T))$, on a $\mathfrak{c}=0$. Par hypothèse, la classe $q^*(\mathfrak{c})$, qui contient un diviseur de la forme $T \times e$, e étant un diviseur sur U, contient aussi un diviseur de la forme $d \times U$, d étant un diviseur sur T; on a donc $T \times e = d \times U + P$, où P est un diviseur principal. Soit t un point de T n'appartenant pas à $\operatorname{Supp} d$, et soit j l'application $x \to (t, x)$ de U dans $T \times U$; alors $j^*(T \times e)$ est défini et égal à e, et $j^*(d \times U)$ est défini et nul; $j^*(P)$ est donc défini et égal à e. Or, comme P est principal, il en est de même de $j^*(P)$, donc de e, d'où $\mathfrak{c}=0$.

Si nous supposons que T est non singulière, l'application $T' \to \mathfrak{M}(T', U)$ (T' ouvert non vide dans T) est, comme on l'a vu, un faisceau. Il résulte immédiatement de ce qu'on vient de dire que $T' \to \mathfrak{M}_0(T', U)$ est un sous-faisceau constant du faisceau $T' \to \mathfrak{M}(T', U)$, isomorphe au faisceau constant de valeur $\mathfrak{G}(U)$. Un faisceau constant étant flasque, on en déduit que l'application $T' \to \mathfrak{N}(T', U)$ est également un faisceau.

Proposition 4. *Soient U une variété semi-complète, T et T' des variétés normales, h un morphisme dominant de T' dans T; les applications h^*: $\mathfrak{M}(T, U) \to \mathfrak{M}(T', U)$ et h^*: $\mathfrak{N}(T, U) \to \mathfrak{N}(T', U)$ sont alors injectives.*

Soit m un élément de $\mathfrak{M}(T, U)$ tel que $h^*(m)=0$; m définit une famille algébrique f de classes de diviseurs de U paramétrée par T. Puisque $h^*(m)=0$, on a $f \circ h=0$; on a donc $f(t)=0$ pour tous les points de la partie dense $h(T')$ de T, d'où $f=0$ (Théorème 2) et par suite $m=0$ (Théorème 3). Supposons maintenant que $h^*(m)$ appartienne au sous-groupe $\mathfrak{M}_0(T', U)$; cela signifie que $f \circ h$ est une application constante de T' dans $\mathfrak{G}(U)$; soit $\mathfrak{c}$ sa valeur. L'application constante $t \to \mathfrak{c}$ est une famille algébrique paramétrée par T, définie par un élément $m_0 \in \mathfrak{M}_0(T, U)$; on a $h^*(m)=h^*(m_0)$, d'où $m=m_0$ et par suite $m \in \mathfrak{M}_0(T, U)$. Ceci démontre la Proposition 4.

Remarque. Si T' est une sous-variété ouverte de T et h l'injection canonique $T' \to T$, on peut donner de la Proposition 4 une démonstration qui ne dépend pas de l'hypothèse que U soit semi-complète. Soit $\mathfrak{k}$ un élément de $\mathfrak{G}(T \times U)$ tel que $h^*(\mathfrak{k})$ soit image réciproque d'un élément de $\mathfrak{G}(T')$ par la projection $T' \times U \to T'$. Si $D \in \mathfrak{k}$, il y a une fonction numérique w' sur $T' \times U$ et un diviseur d' de T' tels que le diviseur induit par D sur $T' \times U$ soit de la forme $\operatorname{div} w' + d' \times U$. La fonction w' se prolonge en une fonction w sur $T \times U$; si $D_1 = D - \operatorname{div} w$, on a

$$\operatorname{Supp} D_1 \subset ((T - T') \cup \operatorname{Supp} d') \times U,$$

d'où il résulte que D_1 est de la forme $d_1 \times U$, d_1 étant un diviseur sur T; ceci

montre que h^* est un homomorphisme injectif de $\mathfrak{M}(T, U)$ dans $\mathfrak{M}(T', U)$. Par ailleurs, si $h^*(\mathfrak{k})$ appartient au noyau de l'homomorphisme $\mathfrak{G}(T' \times U) \to \mathfrak{N}(T', U)$, on voit comme ci-dessus qu'il y a un diviseur principal div w tel que le diviseur induit par $D - \operatorname{div} w$ sur $T' \times U$ soit de la forme $d' \times U + T' \times e$, d' et e étant des diviseurs sur T' et U respectivement; il en résulte que le diviseur induit par $D - \operatorname{div} w - T \times e$ est $d' \times U$, donc que $D - \operatorname{div} w - T \times e$ est de la forme $d \times U$, d étant un diviseur sur T. Ceci montre que h définit une application injective de $\mathfrak{N}(T, U)$ dans $\mathfrak{N}(T', U)$.

On notera que, si g est un morphisme d'une variété U' dans la variété U, g définit un homomorphisme $g^* : \mathfrak{G}(T \times U) \to \mathfrak{G}(T \times U')$ (on désigne encore par g l'application $(t, x') \to (t, g(x'))$ de $T \times U'$ dans $T \times U$). Soient p et p' les projections de $T \times U$ et de $T \times U'$ sur T; il est clair que g^* applique $p^*(\mathfrak{G}(T))$ dans $p'^*(\mathfrak{G}(T))$, donc définit un homomorphisme, encore noté g^*, de $\mathfrak{M}(T, U)$ dans $\mathfrak{M}(T, U')$. Les homomorphismes g^* définissent l'application $U \to \mathfrak{M}(T, U)$ comme foncteur contravariant en son second argument. Si de plus h est un morphisme d'une variété T' dans la variété T, le diagramme

$$\begin{array}{ccc} \mathfrak{M}(T, U) & \xrightarrow{g^*} & \mathfrak{M}(T, U') \\ {\scriptstyle h^*}\downarrow & & \downarrow{\scriptstyle h^*} \\ \mathfrak{M}(T', U) & \xrightarrow{g^*} & \mathfrak{M}(T', U') \end{array}$$

est commutatif; en effet, le morphisme $r : (t', x') \to (h(t'), g(x'))$ définit un homomorphisme de $\mathfrak{G}(T \times U)$ dans $\mathfrak{G}(T' \times U')$ qui applique $p^*(\mathfrak{G}(T))$ dans $p''^*(\mathfrak{G}(T'))$, p'' désignant la projection de $T' \times U'$ sur T' (cela résulte de ce que $p \circ r = h \circ p''$); r définit donc un homomorphisme de $\mathfrak{M}(T, U)$ dans $\mathfrak{M}(T', U')$. Par ailleurs, r est le composé des morphismes $(t', x') \to (h(t'), x')$ et $(t, x') \to (t, g(x'))$, et est aussi le composé des morphismes $(t', x') \to (t', g(x'))$ et $(t', x) \to (h(t'), x)$; notre assertion résulte immédiatement de là. Ce résultat s'exprime en disant que $(T, U) \to \mathfrak{M}(T, U)$ est un bifoncteur, contravariant par rapport à chacun de ses arguments, sur la catégorie des variétés.

De même, si g est un morphisme $U' \to U$, g définit un homomorphisme $g^* : \mathfrak{N}(T, U) \to \mathfrak{N}(T, U')$, et, si h est un morphisme $T' \to T$, le diagramme

$$\begin{array}{ccc} \mathfrak{N}(T, U) & \xrightarrow{g^*} & \mathfrak{N}(T, U') \\ {\scriptstyle h^*}\downarrow & & \downarrow{\scriptstyle h^*} \\ \mathfrak{N}(T', U) & \xrightarrow{g^*} & \mathfrak{N}(T', U') \end{array}$$

est commutatif, comme le lecteur le vérifiera immédiatement; le symbole $\mathfrak{N}$ apparait donc comme un symbole de bifoncteur sur la catégorie des variétés.

Par ailleurs, nous avons déjà observé que, si T et U sont des variétés, l'isomorphisme canonique $U \times T \to T \times U$ définit un isomorphisme $\sigma_{T,U}$: $\mathfrak{N}(T,U) \to \mathfrak{N}(U,T)$. Soient g un morphisme d'une variété U' dans la variété U et h un morphisme d'une variété T' dans la variété T; soient r le morphisme $(t',x') \to (h(t'),g(x'))$ de $T' \times U'$ dans $T \times U$ et r' le morphisme $(x',t) \to (g(x'),h(t'))$ de $U' \times T'$ dans $U \times T$; r et r' définissent des homomorphismes r^*: $\mathfrak{N}(T,U) \to \mathfrak{N}(T',U')$, r'^*: $\mathfrak{N}(U,T) \to \mathfrak{N}(U',T')$. On vérifie immédiatement que le diagramme

$$\begin{array}{ccc} \mathfrak{N}(T,U) & \xrightarrow{\sigma_{T,U}} & \mathfrak{N}(U,T) \\ {\scriptstyle r^*}\downarrow & & \downarrow{\scriptstyle r'^*} \\ \mathfrak{N}(T',U') & \xrightarrow{\sigma_{T',U'}} & \mathfrak{N}(U',T') \end{array}$$

est commutatif.

PROPOSITION 5. *Soient T une sous-variété ouverte de la droite projective $\bar{K}$ et U une variété normale; $\mathfrak{N}(T,U)$ se réduit alors à $\{0\}$.*

Soit U_1 l'ensemble des points simples de U; il existe alors un homomorphisme injectif $\mathfrak{N}(U,T) \to \mathfrak{N}(U_1,T)$ (cf. la remarque qui suit la Proposition 4); tenant compte des isomorphismes $\mathfrak{N}(U,T) \cong \mathfrak{N}(T,U)$, $\mathfrak{N}(U_1,T) \cong \mathfrak{N}(T,U_1)$, on voit qu'il de démontrer la Proposition 5 dans le cas où U est non-singulière. Supposons qu'il en soit ainsi; $\bar{K} \times U$ est alors non-singulière; l'injection canonique $T \times U \to \bar{K} \times U$ définit donc une application surjective de $\mathfrak{D}(\bar{K} \times U)$ sur $\mathfrak{D}(T \times U)$ (Corollaire à la Proposition 1, § I), et par suite aussi une application surjective de $\mathfrak{N}(\bar{K},U)$ sur $\mathfrak{N}(T,U)$. Il suffira donc de montrer que $\mathfrak{N}(\bar{K},U) = \{0\}$, ou encore que $\mathfrak{N}(U,\bar{K}) = \{0\}$. Comme U est normale et $\bar{K}$ complète, les éléments de $\mathfrak{M}(U,\bar{K})$ sont en correspondence bi-univoque avec les applications algébriques de U dans $\mathfrak{G}(\bar{K})$. Or la structure du groupe $\mathfrak{G}(\bar{K})$ est bien connue: une condition nécessaire et suffisante pour que deux diviseurs sur $\bar{K}$ soient équivalents est qu'ils aient même degré. Si donc on appelle degré d'une classe $\mathfrak{c} \in \mathfrak{G}(\bar{K})$ le degré commun des diviseurs de $\mathfrak{c}$, pour montrer que $\mathfrak{N}(U,\bar{K}) = \{0\}$, il suffira de montrer que, si f est une application algébrique de U dans $\mathfrak{G}(\bar{K})$, le degré de $f(x)$ $(x \in U)$ ne dépend pas de x (car cela entraînera que f est constante). En vertu du caractère connexe de U, il suffira de montrer que tout point de U admet un voisinage U_0 tel que le degré de $f(x)$ reste constant pour $x \in U_0$. Or il y a un voisinage U_0 de x et une famille algébrique f de diviseurs de $\bar{K}$ paramétrée par U_0 tels

que $f(x) = \mathrm{Cl}.\bar{f}(x)$ si $x \in U_0$ (Proposition 2), et le degré de $\bar{f}(x)$ reste constant pour $x \in U_0$ (Proposition 3, § II) ; la Proposition 5 est donc établie.

V. Criteres de rationalité (II).

PROPOSITION 1. *Soient T une variété et (T', h) un revêtement séparable de T. Soit f une application de T dans l'ensemble des classes de diviseurs d'une variété semi-complète U. Si $f \circ h$ est une application algébrique de T' dans $\mathfrak{G}(U)$, il y a une partie ouverte non vide T_0 de T telle que la restriction de f à T_0 soit une application algébrique de T_0 dans $\mathfrak{G}(U)$.*

On peut évidemment supposer T normale. Il existe un revêtement (T'_1, h_1) de T' tel que $(T'_1, h \circ h_1)$ soit revêtement galoisien normal de T; comme $f \circ h \circ h_1$ est une application algébrique de T'_1 dans $\mathfrak{G}(U)$, on voit qu'on peut supposer que (T', h) est galoisien. Nous désignerons par G le groupe des automorphismes du revêtement (T', h) ; nous considérerons G comme opérant aussi sur $T' \times U$. L'application $f \circ h$ est définie par une classe $\mathfrak{k}'$ de diviseurs de $T' \times U$; nous choisirons un point $x_1 \in U$ et un représentant D' de $\mathfrak{k}'$ tel que $T' \times \{x_1\} \not\subset \mathrm{Supp}\, D'$. Si s est un élément de G, on a $f \circ h \circ s = f \circ h$; il en résulte que $s^*(\mathfrak{k}') - \mathfrak{k}'$ est image réciproque d'un élément de $\mathfrak{G}(T')$ par la projection $T' \times U$ sur T', donc que $s^*(D') - D'$ est de la forme $\mathrm{div}\, w_s' + d_s' \times U$, où w_s' est une fonction numérique sur $T' \times U$ et d_s' un diviseur sur T'. Désignons par k l'application $t' \to (t', x_1)$ de T' dans $T' \times U$; montrons que l'on peut choisir w_s' de telle manière que w_s' soit composable avec k et que $w_s' \odot k = 1$. Puisque $T' \times \{x_1\} \not\subset \mathrm{Supp}\, D'$, $k^*(D')$ est défini; il en est de même de $k^*(d' \times U) = d'$; par suite, $k^*(\mathrm{div}\, w_s')$ est défini, ce qui montre que w_s' est composable avec k. Soit p' la projection $T' \times U \to T'$; soit $z' = w_s' \odot k \odot p'$, d'où $z' \odot k = w_s' \odot k$; on a $s^*(D') - D' = \mathrm{div}\, z'^{-1} w_s' + \mathrm{div}\, z' + d_s' \times U$; or on a $\mathrm{div}\, z' = \mathrm{div}(w_s' \odot k) \times U$; remplaçant w_s' par $z^{-1} w_s'$ et d_s' par $d_s' + \mathrm{div}(w_s' \odot k)$, on voit qu'on peut supposer que $w_s' \odot k = 1$. La fonction w_s' est alors uniquement determinée. Pour le montrer, nous avons à établir que, si w'' est une fonction numérique sur $T' \times U$ telle que $\mathrm{div}\, w''$ soit de la forme $d'' \times U$, d'' étant un diviseur sur T', et si $w'' \odot k = 1$, on a $w'' = 1$. Pour tout $t' \in T'$, soit $j'_{t'}$ l'application $x \to (t', x)$ de U dans $T' \times U$; si $t' \notin \mathrm{Supp}\, d''$, $j_t^*(\mathrm{div}\, w'')$ est défini et nul, ce qui montre que $w'' \odot j'_{t'}$ est une fonction de diviseur nul sur U, et par suite constante, puisque U est semi-complète. Par ailleurs, (t', x_1) n'est pas dans $\mathrm{Supp}\,\mathrm{div}\, w''$, de sorte que w'' est définie en (t', x_1) ; comme $w'' \odot k = 1$, on a $w''(t', x_1) = 1$, d'où $w'' \odot j'_{t'} = 1$. Ceci étant vrai pour tous les points de la partie ouverte non vide $T' - \mathrm{Sup}\, d''$ de T', il en résulte immédiatement que $w'' = 1$.

Ceci étant, on a, $s, t \in G$, $(st)^*(D') - D' = t^*(s^*(D') - D') + t^*(D') - D'$; comme $((w_s' \odot t)w_t') \odot k = 1$, il résulte de l'assertion d'unicité que nous venons de faire que l'on a $w_{st}' = (w_s' \odot t)w_t'$. Le groupe G opère à droite sur le corps des fonctions numériques sur $T' \times U$ au moyen des applications $w' \to w' \odot s$; la formule précédente signifie que l'application $w' \to w_s'$ est un cocycle pour G à valeurs dans le corps des fonctions sur $T' \times U$. Il est bien connu qu'il existe alors une fonction numérique $w' \neq 0$ sur $T' \times U$ telle que $w_s' = w'^{-1}(w' \odot s)$ pour tout $s \in G$. Si donc on pose $D'_1 = D' - \operatorname{div} w'$, on a $s^*(D'_1) - D'_1 = d_s' \times U$. Soit T_1 une partie ouverte non vide de T qui ne rencontre aucun des ensembles $h(\operatorname{Supp} d_s')$, et soit $T_1' = h^{-1}(T_1)$; on peut encore considérer G comme opérant sur $T_1' \times U$. Si D''_1 est le diviseur induit par D'_1 sur $T_1' \times U$, on a $s^*(D''_1) = D''_1$ pour tout $s \in G$. Il existe une partie ouverte non vide T_0 de T_1 telle que (T', h) soit non ramifié en tout point de $h^{-1}(T_0)$. Si h_0 est la restriction de h à $h^{-1}(T_0)$ et r le morphisme $(t', x) \to (h_0(t'), x)$ de $h^{-1}(T_0) \times U$ dans $T_0 \times U$, $(h^{-1}(T_0) \times U, r)$ est un revêtement galoisien non ramifié de $T_0 \times U$. Si D''_0 est le diviseur induit par D''_1 sur $h^{-1}(T_0) \times U$, il résulte de la Proposition 3, § I que D''_0 se met sous la forme $h^*(D_0)$, D_0 étant un diviseur sur $T_0 \times U$. Soit $\mathfrak{k}_0$ la classe de D_0 dans $\mathfrak{G}(T_0 \times U)$; si i est l'injection canonique de $T_0 \times U$ dans $T \times U$, et h_0 la restriction de h à $h^{-1}(T_0)$, $h_0^*(\mathfrak{k}_0)$ est évidemment égal à $i^*(\mathfrak{k}')$. Si donc f_0 est la famille algébrique de diviseurs de U paramétrée par T_0 définie par $\mathfrak{k}_0$, $f_0 \circ h_0$ est la restriction de $f \circ h$ à $h^{-1}(T_0)$; comme h est surjectif, f_0 est la restriction de f à T_0; cette dernière est donc une famille algébrique.

PROPOSITION 2. *Soient U une variété semi-complète, G un groupe algébrique, H un sous-groupe fermé de G, h l'application canonique de G sur G/H. Soit g un homomorphisme de G dans le groupe des classes de diviseurs d'une variété normale et complète U; supposons que H soit contenu dans le noyau de g et que g soit une famille algébrique paramétrée par G; alors l'application f de G/H dans $\mathfrak{G}(U)$ telle $f \circ h = g$ est une famille algébrique paramétrée par G/H.*

Soient e l'élément neutre de G et H_0 la composante connexe de e dans H. Comme e est simple sur G et H_0, il existe un système $(u_1, \cdots, u_n)$ de variables uniformisantes sur G en e, prenant en ce point la valeur 0, tel que l'idéal de définition en e de la variété H_0 soit engendré par $u_1, \cdots, u_m$, m étant un entier $\leqq n$. L'idéal engendré par $u_{m+1}, \cdots, u_n$ dans l'anneau local de e est l'idéal de définition en e d'une sous-variété fermée T' de G; T' est de dimension $n - m$, e est un point isolé de $T' \cap H$, e est simple sur T' et l'espace tangent à G en e est somme directe des espaces tangents à H_0 et à

T' en e. L'application h induit un morphisme h' de T' dans G/H; comme e est isolé dans $h'^{-1}(h(e)) = T' \cap H$, on a $\dim h'(T') = \dim T' = n - m = \dim G/H$; h' est donc dominant; c'est un morphisme de degré fini. Nous allons voir qu'il est séparable. Il suffira évidemment pour cela de démontrer que le morphisme $(t', s) \to h'(t')$ $(s \in H_0)$ de $T' \times H_0$ dans G/H est séparable. Or ce morphisme est composé de l'application $(t', s) \to t's$ de $T \times H_0$ dans G et de l'application h de G sur G/H. Comme h est un morphisme séparable, il suffira de montrer que le morphisme $(t', s) \to t's$ de $T' \times H_0$ dans G est séparable (des considérations de dimension montrent tout de suite que ce morphisme est dominant). Il suffira pour cela de montrer que ce morphisme, soit θ, est non ramifié en (e, e). Or l'image par l'application dérivée $D\theta$ de θ en (e, e) de l'espace tangent à $T' \times H_0$ en (e, e) contient les espace tangents à T' et à H_0 en e, puisque $\theta(t', e) = t'$ $(t' \in T')$, $\theta(e, s) = s$ $(s \in H)$; cette image est donc l'espace tangent à G en e tout entier. Or l'espace tangent à $T' \times H_0$ en (e, e) est de dimension n égale à la dimension de l'espace tangent en e; $D\theta$ est donc injectif, ce qui démontre notre assertion.

Puisque h' est un morphisme dominant de degré fini, il y a une partie ouverte non vide W_0 de G/H telle que, h'_0 désignant la restriction de h' à $h'^{-1}(W_0) = T'_0$, (T'_0, h'_0) soit un revêtement de W_0 ([2], Proposition 4, chap. IV, §III). L'application $f \circ h'_0$ est la restriction de g à T'_0 et est par suite une application algébrique de T'_0 dans $\mathfrak{G}(U)$. Faisant usage de la Proposition 1, on voit qu'il existe une partie ouverte non vide W_1 de W_0 telle que la restriction de f à W_1 soit une application algébrique de W_1 dans $\mathfrak{G}(U)$. Soit z_1 un point de W_1; si z_2 est un point quelconque de G/H, il y a une opération s de G qui transforme z_1 en z_2; si $z \in W_1$, on a $f(s \cdot z) = g(s) + f(z)$, comme il résulte du fait que g est un homomorphisme. Il en résulte que la restriction de f au voisinage $s \cdot W_1$ de z_2 est une application algébrique de ce voisinage dans $\mathfrak{G}(U)$. Or G/H est une variété non singulière; il résulte alors de la Proposition 3, § IV que f est une famille algébrique de classes de diviseurs.

Chapitre II.

I. Construction de la jacobienne. Soit C une courbe normale et complète. Pour tout diviseur d sur C, désignons par $\delta(d)$ le degré de d et par $\lambda(d)$ la dimension de l'espace vectoriel composée des fonctions sur C qui sont multiples de d. De la théorie des courbes, nous n'utiliserons que les résultats suivants: 1) le degré de tout diviseur principal est 0; 2) les nombres $\delta(d) + \lambda(d)$, pour tous les diviseurs d de C, forment un ensemble borné

inférieurement; si $-g+1$ est le plus petit de ces nombres, g est un entier $\geqq 0$ qu'on appelle le *genre* de la courbe C.

Soient d un diviseur et V l'espace des fonctions qui sont multiples de d. Si $x_1, \cdots, x_m$ sont des points de C, l'espace V' des fonctions qui sont multiples de $d+\sum_{i=1}^{m} x_i$ est de codimension $\leqq m$ dans V. Procédant par récurrence sur m, on voit qu'il suffit de le montrer dans le cas où $m=1$. Soit alors t une fonction de définition de d en x_1; si $u \in V$, $t^{-1}u$ est définie en x_1, et une condition nécessaire et suffisante pour que u soit multiple de x_1+d est que $(t^{-1}u)(x_1)=0$; ceci démontre notre assertion, $u \to t^{-1}u(x_1)$ étant une forme linéaire sur V.

Si $\mathfrak{k}$ est une classe de diviseurs de C, tous les diviseurs de $\mathfrak{k}$ ont le même degré qu'on note $\delta(\mathfrak{k})$ et qu'on appelle le *degré de la classe* $\mathfrak{k}$. Par ailleurs, si d et d' sont des diviseurs de $\mathfrak{k}$, on a $\lambda(d)=\lambda(d')$, car, si V et V' sont les espaces de fonctions multiples de d et d' respectivement, et si u est une fonction de diviseur $d'-d$, on a $V'=uV$.

Toute classe $\mathfrak{k}$ de degré $r \geqq g$ contient un diviseur positif. Soit en effet d un diviseur de la classe $\mathfrak{k}$; on a

$$\lambda(-d) \geqq -\delta(-d)+1-g=r+1-g>0,$$

ce qui montre qu'il existe une fonction $u \neq 0$ telle que $\operatorname{div} u+d \geqq 0$.

Pour tout entier $m \leqq 1-g$, il existe un diviseur d de degré m tel que $\lambda(d)+\delta(d)=1-g$. Il suffit en effet de montrer que, si l'assertion est vraie pour m, elle l'est aussi pour $m-1$, et, si $m<1-g$, pour $m+1$. Soit d un diviseur de degré m tel que $\lambda(d)+\delta(d)=1-g$. Si x est un point quelconque de C, on a $\delta(d-1\cdot x)=\delta(d)-1$, $\lambda(d) \geqq \lambda(d-1\cdot x)-1$, et par suite $\lambda(d-1\cdot x)+\delta(d-1\cdot x) \leqq 1-g$; comme on a aussi $\lambda(d-1\cdot x)+\delta(d-1\cdot x) \geqq 1-g$, on voit que l'assertion est vraie pour $m-1$. Supposons maintenant que $m<1-g$; on a alors $\lambda(d)>0$. Soit u une fonction $\neq 0$ qui est multiple de d; il est clair qu'il y a un point $x \in C$ tel que u ne soit pas multiple de $d+1\cdot x$; on a donc alors $\lambda(d+1\cdot x)=\lambda(d)-1$, et par suite $\lambda(d+1\cdot x)+\delta(d+1\cdot x)=1-g$, ce qui montre que l'assertion est vraie pour $m+1$.

On va déduire de là qu'il existe une classe de degré g qui ne contient qu'un seul diviseur $\geqq 0$. Soit d un diviseur de degré $-g$ tel que $\lambda(d)+\delta(d)=1-g$, d'où $\lambda(d)=1$. Soit u_1 une base de l'espace des fonctions multiples de d. Si d' est un diviseur positif de la classe de $-d$, on a $d'=\operatorname{div} u-d$, u étant une fonction numérique; comme d' est $\geqq 0$, u est multiple de d, d'où $u=cu_1$, c étant un élément $\neq 0$ de k. Il en résulte bien que la classe de $-d$ ne contient qu'un diviseur positif.

On appelle *non spéciales* les classes de degré g qui ne contiennent qu'on seul diviseur entier; nous dirons qu'un diviseur entier de degré g est non spécial si sa classes est non spéciale.

Soit S^g la puissance symétrique g-ième de la courbe C, et soit d_g la famille canonique de diviseurs de C paramétrée par S^g. *Nous désignerons par* Ω *l'ensemble des points* $z \in S^g$ *tels que* $d_g(z)$ *soit non spécial; de plus, pour tout* $z \in S^g$, *nous désignerons par* $\mathfrak{d}_g(z)$ *la classe de* $d_g(z)$.

PROPOSITION 1. *L'ensemble* Ω *est ouvert dans* S^g.

Disons qu'on diviseur est spécial s'il n'est pas non spécial. Si $d_g(z)$ est spécial, on a $\lambda(-d_g(z)) > 1$, d'où, si $x \in C$, $\lambda(-d_g(z) + 1 \cdot x) > 0$. Il y a donc un diviseur entier de degré g, donc de la forme $d_g(z')$, avec $z' \in S^g$, qui est dans la classe de $d_g(z)$ et qui est $\geqq 1 \cdot x$. Réciproquement, si, pour tout $x \in C$, $\mathfrak{d}_g(z)$ contient un diviseur entier $\geqq 1 \cdot x$, il est clair que $\mathfrak{d}_g(z)$ est spéciale. Il suffira donc de montrer que, si $x \in C$, l'ensemble B_x des $z \in S^g$ tels que $\mathfrak{d}_g(z)$ contienne un diviseur entier $\geqq 1 \cdot x$ est fermé. Or, l'ensemble A_x des $z' \in S^g$ tels que $d_g(z') - 1 \cdot x \geqq 0$ est fermé (Corollaire 3 au Théorème 1, I, § II). Par ailleurs, l'ensemble H des couples $(z', z) \in S^g \times S^g$ tels que $\mathfrak{d}_g(z') - \mathfrak{d}_g(z) = 0$ est fermé (Théorème 2, I, § IV); B_x est l'image de $H \cap (A_x \times S^g)$ par la projection de $S^g \times S^g$ sur son second facteur, et est par suite fermé puisque S^g est une variété complète.

PROPOSITION 2. *Soit f une famille algébrique de classes de diviseurs de degré g de C paramétrée par une variété T. L'ensemble T_1 des $t \in T$ tels que $f(t)$ soit non spécial est ouvert. Supposons que T soit normale, et qu'il existe un point $t_0 \in T$ tel que $f(t_0)$ soit une classe non spéciale et contienne un diviseur entier de degré g qui soit somme de g points distincts de C. Il existe alors un morphisme φ de T_1 dans Ω tel que $f(t) = \mathfrak{d}_g(\varphi(t))$ pour tout $t \in T_1$.*

Si $t \in T$, la classe $f(t)$, qui est de degré g, contient un diviseur entier, donc de la forme $d_g(z)$, $z \in S^g$. L'ensemble H des $(t, z) \in T \times S^g$ tels que $f(t) - \mathfrak{d}_g(z) = 0$ est fermé (Théorème 2, I, § IV); la projection de $T \times S^g$ induit une application surjective de H sur T. L'ensemble $H \cap (T \times (S^g - \Omega))$ est fermé; il en est de même de son image par la projection $T \times S^g \to T$, puisque S^g est complète. Or, cette image est $T - T_1$; en effet, pour que $t \in T_1$, il suffit qu'il existe un point $z \in \Omega$ tel que $(t, z) \in H$, et, s'il en est ainsi, z est le seul point de S^g tel que $(t, z) \in H$, comme il résulte du fait que $f(t)$ ne contient qu'un seul diviseur entier. Il résulte de là que T_1 est ouvert. Soit $H_1 = H \cap (T_1 \times S^g)$; H_1 est relativement fermé dans $T_1 \times S^g$, et la projection $T_1 \times S^g \to T_1$ induit une bijection p_1 de H_1 sur T_1. Il en résulte d'abord que,

si $H_1 \neq \emptyset$, H_1 est une sous-variété de $T_1 \times S^g$. Il existe en effet une composante irréductible H_1' de H_1 telle que $p_1(H_1')$ soit dense dans T_1. Mais H_1' est fermé dans $T_1 \times S^g$; S^g étant complète, $p_1(H_1')$ est fermé dans T_1, donc égal à T_1; p_1 étant bijectif, il en résulte que $H_1' = H_1$.

Observons maintenant que l'ensemble des points $z \in S^g$ tels que $d_g(z)$ soit somme de g points distincts est ouvert. Son complémentaire est en effet l'image par l'application canonique s_g: $C^g \to S^g$ de l'ensemble des points $(x_1, \cdots, x_g)$ de C^g tels que l'on ait $x_i = x_j$ pour au moins un couple d'indices distincts i et j, ensemble qui est évidemment fermé. Soit Ω_1 l'ensemble des points $z \in \Omega$ tels que $d_g(z)$ soit somme de g points distincts; il est ouvert. Nous supposons que T est normale et que $H_1 \cap (T_1 \times \Omega_1)$ n'est pas vide; nous allons alors montrer que p_1 est un isomorphisme de H_1 sur T_1. L'ensemble $H_1 \cap (T_1 \times \Omega_1)$ est une sous-variété ouverte de H_1; l'ensemble des points en lesquels cette variété est normale est lui-même une sous-variété ouverte H_2 de H_1. Par ailleurs, il existe une sous-variété ouverte T_2 de T_1 et une famille algébrique $\bar{f}$ de diviseurs de C paramétrée par T_2 telles que, pour $t \in T_2$, $f(t)$ soit la classe de $\bar{f}(t)$. Comme p_1 est surjectif, $H_3 = H_2 \cap (T_2 \times S^g)$ est une sous-variété ouverte de H_2. La formule $m(t,z) = d_g(z) - \bar{f}(t)$ $((t,z) \in H_3)$ définit une famille algébrique m de diviseurs de C paramétrée par H_3; on a $m(t,z) \sim 0$ pour tout $(t,z) \in H_3$. Soit M le diviseur de definition de m; comme H_3 est normale et C complète, M est de la forme $\operatorname{div} w + e \times C$, w étant une fonction numérique $H_3 \times C$ et e étant un diviseur sur H_3 (Théorème 3, I, § IV). Soit $H_4 = H_3 - \operatorname{Supp} e$; H_4 est une sous-variété ouverte de H_3, et le diviseur de définition de la restriction m_4 de m à H_4 est principal. Soit (t_1, z_1) un point de H_4; nous allons montrer que p_1 est non ramifié en ce point. Soit Λ un vecteur tangent à H_4 (donc aussi à H_1, ou à $T_1 \times S^g$) en (t_1, z_1) dont l'image par la dérivée de p_1 soit nulle; soit L l'image de Λ par la restriction q_4 de la projection $T_1 \times S^g \to S^g$ à H_4. Pour montrer que $\Lambda = 0$, il suffira de montrer que $L = 0$. Le point z_1 appartient à Ω_1, de sorte que $d_g(z_1)$ est somme de g points distincts de C; la famille d_g est donc infinitésimalement injective en z_1 (I, § III); pour montrer que $L = 0$, il suffira donc de montrer que $< L, d_g > = 0$. Or, si p_4 est la restriction p_1 à H_4, on a $m_4 = d_g \circ q_4 - f \circ p_4$; comme l'image de Λ par la dérivée de p_4 est nulle, on a $< \Lambda, f \circ p_4 > = 0$ (Lemme 1, I, § III), d'où $< \Lambda, m_4 > = < \Lambda, d_g \circ p_4 >$ $= < L, d_g >$. Puisque le diviseur de définition de m_4 est principal, le diviseur additif $< \Lambda, m_4 >$ est principal; il est représenté par une fonction numérique s sur C, que l'on peut supposer $\neq 0$ (car $< \Lambda, m_4 >$ est aussi représenté par $s + 1$). Par ailleurs, d_g est une famille de diviseurs $\geqq 0$; son diviseur de définition est donc $\geqq 0$. Le diviseur additif $< d_g, L >$ est représenté par s;

si donc u est une fonction de définition de $d_g(z_1)$ en un point x quelconque de C, su est définie en x (Lemme 2, I, § III). Il en résulte que $\operatorname{div} s + d_g(z_1) \geqq 0$; or ce diviseur appartient à $\mathfrak{d}_g(z_1)$; mais, comme $z_1 \in \Omega$, $\mathfrak{d}_g(z_1)$ ne contient qu'un seul diviseur $\geqq 0$, à savoir $d_g(z_1)$; on a donc $\operatorname{div} s = 0$. Comme C est complète, s est une constante; comme s est une fonction de définition du diviseur additif $< d_g, L >$, ce dernier est nul, d'où $L = 0$.

Comme p_1 est un morphisme dominant non ramifié en (t_1, z_1), ce morphisme est séparable ([2], Corollaire 2 à la Proposition 3, II, chap. VI). Comme p_1 est injectif, il est aussi radiciel; p_1 est donc birationnel. Comme T_1 est une variété normale, il résulte du théorème principal de Zariski que p_1 est un isomorphisme de H_1 sur T_1. En composant l'isomorphisme réciproque de p_1 avec la restriction à H_1 de la projection $T \times S^g \to S^g$, on obtient un morphisme φ de T dans S^g; il est clair que $f(t) = \mathfrak{d}_g(\varphi(t))$ pour $t \in T_1$.

Nous désignerons par $\mathfrak{G}_0(C)$ (resp. $\mathfrak{G}_g(C)$) l'ensemble des classes de diviseurs de degré 0 (resp. g) de C. Si $\mathfrak{c} \in \mathfrak{G}_g(C)$, nous désignerons par $\theta_\mathfrak{c}$ l'application $z \to \mathfrak{d}_g(z) - \mathfrak{c}$ de S^g dans $\mathfrak{G}_0(C)$, par $V(\mathfrak{c})$ l'ensemble $\theta_\mathfrak{c}(\Omega)$ et par $\theta_\mathfrak{c}'$ la restriction de $\theta_\mathfrak{c}$ à Ω. Si $z \in \Omega$, $\theta_\mathfrak{c}^{-1}(\theta_\mathfrak{c}(z))$ se compose du seul point z, puisque $\mathfrak{d}_g(z)$ ne contient qu'un seul diviseur entier (et que l'application $d_g \colon S^g \to \mathfrak{D}(C)$ est injective). Et particulier, $\theta_\mathfrak{c}'$ est une bijection de Ω sur $V(\mathfrak{c})$; elle permet de transporter à $V(\mathfrak{c})$ la structure de variété de Ω. L'ensemble $\mathfrak{G}_0(C)$ est la réunion des ensembles $V(\mathfrak{c})$ pour tous les $\mathfrak{c} \in \mathfrak{G}_g(C)$; en effet, si $\mathfrak{k}$ est un élément de $\mathfrak{G}_0(C)$, $\mathfrak{c} \to \mathfrak{k} + \mathfrak{c}$ est une permutation de l'ensemble $\mathfrak{G}_g(C)$, de sorte que, si $z_1 \in \Omega$, il existe toujours un $\mathfrak{c} \in \mathfrak{G}_g(C)$ tel que $\mathfrak{k} + \mathfrak{c} = \mathfrak{d}_g(z_1)$, d'où $\mathfrak{k} \in V(\mathfrak{c})$. On voit en même temps que, si z_1 est un point quelconque de S^g, $\mathfrak{c} \to \mathfrak{d}_g(z_1) - \mathfrak{c}$ est une bijection de $\mathfrak{G}_g(C)$ sur $\mathfrak{G}_0(C)$; or toute classe de degré g contient un diviseur entier, donc de la forme $\mathfrak{d}_g(z)$, $z \in S^g$; comme $\mathfrak{k} \to -\mathfrak{k}$ est une permutation de $\mathfrak{G}_0(C)$, on voit que, pour tout $z_1 \in S^g$, l'application $z \to \mathfrak{d}_g(z_1)$ est une surjection de S^g sur $\mathfrak{G}_g(C)$. Nous allons maintenant voir qu'il existe sur $\mathfrak{G}_0(C)$ une structure de variété (et, naturellement, une seule) telle que, pour tout $\mathfrak{c} \in \mathfrak{G}_g(C)$, $V(\mathfrak{c})$ soit une sous-variété ouverte de cette variété. Pour montrer qu'il en est ainsi, il suffit de montrer que les conditions suivantes sont satisfaites: 1) si $\mathfrak{c}, \mathfrak{c}' \in \mathfrak{G}_g(C)$, $V(\mathfrak{c}) \cap V(\mathfrak{c}')$ est ouvert dans $V(\mathfrak{c})$ et $V(\mathfrak{c}')$ et les structures de variété induites sur cet ensemble par celles de $V(\mathfrak{c})$ et de $V(\mathfrak{c}')$ sont identiques; de plus, l'ensemble des points de la forme $(\mathfrak{k}, \mathfrak{k})$, $\mathfrak{k} \in V(\mathfrak{c}) \cap V(\mathfrak{c}')$, est fermé dans $V(\mathfrak{c}) \times V(\mathfrak{c}')$; 2) l'ensemble $\mathfrak{G}_0(C)$ est la réunion d'un nombre fini des ensembles $V(\mathfrak{c})$.

Vérifions d'abord la condition 1; $V(\mathfrak{c}) \cap V(\mathfrak{c}')$ est l'image par $\theta_\mathfrak{c}$ (resp. $\theta_{\mathfrak{c}'}$) d'une partie Ω' (resp. Ω'') de Ω. L'ensemble Ω' est l'ensemble des $z' \in \Omega$

tels que $\mathfrak{d}_g(z') - \mathfrak{c} + \mathfrak{c}'$ soit non spéciale. Or, l'application

$$f\colon z' \to \mathfrak{d}_g(z') - \mathfrak{c} + \mathfrak{c}' \qquad (z' \in \Omega)$$

est une famille algébrique de classes de diviseurs de C paramétrée par Ω; ceci montre que Ω' est ouvert (Proposition 2); on voit de même que Ω'' est ouvert; $V(\mathfrak{c}) \cap V(\mathfrak{c}')$ est donc ouvert dans $V(\mathfrak{c})$ et $V(\mathfrak{c}')$. Montrons qu'il existe un point $z_0 \in \Omega$ tel que $f(z_0)$ soit somme de g points distincts de C. Considérons pour cela l'ensemble H des points $(z', z'') \in \Omega \times S^g$ tels que l'on ait $\mathfrak{d}_g(z') = \mathfrak{d}_g(z'') + \mathfrak{c} - \mathfrak{c}'$. Cet ensemble est fermé; si p_1 et p_2 sont les restrictions à H des projections de $\Omega \times S^g$ sur son premier et son second facteur, p_1 est surjectif (car, si $z' \in \Omega$, $\mathfrak{d}_g(z')$ peut se mettre sous la forme $\mathfrak{d}_g(z'') + \mathfrak{c} - \mathfrak{c}'$, avec un $z'' \in S^g$) et p_2 est injectif, car, si $\mathfrak{d}_g(z'') + \mathfrak{c} - \mathfrak{c}'$ est non spéciale, cette classe ne contient qu'un diviseur entier. On déduit de la première assertion que $\dim H \geqq g$, de la seconde que $\dim H = \dim p_2(H) \leqq g$; on a donc $\dim p_2(H) = \dim H = g$, et $p_2(H)$ est dense dans S^g, donc rencontre l'ensemble Ω_1 des points $z'' \in \Omega$ tels que $d_g(z'')$ soit somme de g points distincts. Il résulte alors de la Proposition 2 qu'il existe un morphisme ω de Ω' dans Ω'' tel que $f(z) = \mathfrak{d}_g(\omega(z))$ $(z \in \Omega')$. On voir de même qu'il existe un morphisme ω' de Ω'' dans Ω' tel que $\mathfrak{d}_g(z'') + \mathfrak{c}' - \mathfrak{c} = \mathfrak{d}_g(\omega'(z''))$ $(z'' \in \Omega'')$. Il s'en suit que ω et ω' sont des isomorphismes réciproques l'un de l'autre. Comme ω est l'application définie par la condition que $\theta_{\mathfrak{c}'} \circ \omega = \theta_{\mathfrak{c}}$, on voit que $V(\mathfrak{c})$ et $V(\mathfrak{c}')$ induisent la même structure de variété sur leur intersection. L'ensemble des $(\mathfrak{k}, \mathfrak{k})$, $\mathfrak{k} \in V(\mathfrak{c}) \times V(\mathfrak{c}')$, est l'image part l'application $(z', z'') \to (\theta_{\mathfrak{c}}(z'), \theta_{\mathfrak{c}'}(z''))$ de l'ensemble $H \cap (\Omega \times \Omega)$, qui est fermé dans $\Omega \times \Omega$; il est donc fermé dans $V(\mathfrak{c}) \times V(\mathfrak{c}')$.

Il reste à montrer que $\mathfrak{G}_0(C)$ est la réunion d'un nombre fini des variétés $V(\mathfrak{c})$. Soit $\mathfrak{c}_0$ une classe quelconque de degré g; $\mathfrak{G}_0(C)$ est alors l'image de S^g par l'application $\theta_{\mathfrak{c}_0}\colon z \to \mathfrak{d}_g(z) - \mathfrak{c}_0$; pour tout $\mathfrak{c} \in \mathfrak{G}_g(C)$, $V(\mathfrak{c})$ est l'image par $\theta_{\mathfrak{c}_0}$ de l'ensemble $W(\mathfrak{c})$ des $z \in S^g$ tels que $\mathfrak{d}_g(z) + \mathfrak{c} - \mathfrak{c}_0$ soit non spéciale, ensemble qui est ouvert en vertu de la Proposition 2. La variété S^g est la réunion des ensembles ouverts $W(\mathfrak{c})$; elle est donc la réunion d'un nombre fini d'entre eux, ce qui démontre notre assertion.

La variété J dont l'ensemble de points est $\mathfrak{G}_0(C)$ et dont les $V(\mathfrak{c})$ sont des sous-variétés ouvertes s'appelle la *jacobienne* de C.

THÉORÈME 1. *Soient C une courbe normale et complète, T une variété normale, t_0 un point de T, f une application de T dans $\mathfrak{G}(C)$, J la jacobienne de C. Pour que f soit une famille algébrique de classes de diviseurs de C paramétrée par T, il faut et suffit que l'application $t \to f(t) - f(t_0)$ soit un morphisme de T dans J.*

Supposons d'abord que f soit algébrique. On sait que le degré de $f(t)$ est indépendant de t (Proposition 3, I, II) ; il en résulte que $f(t) - f(t_0) \in J$ pour tout $t \in T$. Soit t_1 un point de T. Soient $\mathfrak{k}_0$ un élément quelconque de $\mathfrak{G}_0(C)$ et $\mathfrak{c}_1$ un élément de $\mathfrak{G}_g(C)$ tel que $f(t_1) - f(t_0) + \mathfrak{k}_0 + \mathfrak{c}_1$ contienne un diviseur entier non spécial qui soit la somme de g points distincts de C. Il résulte alors de la Proposition 2 que l'ensemble T_1 des $t \in T$ tels que $f(t) - f(t_0) + \mathfrak{k}_0 + \mathfrak{c}_1$ soit non spéciale est un voisinage ouvert de t_1 et qu'il existe un morphisme φ de T_1 dans Ω tel que

$$f(t) - f(t_0) + \mathfrak{k}_0 + \mathfrak{c}_1 = \mathfrak{d}_g(\varphi(t))$$

pour $t \in T_1$; la restriction à T_1 de l'application $t \to f(t) - f(t_0)$ est donc l'application $t \to \mathfrak{d}_g(\varphi(t)) - \mathfrak{c}_1$ qui est un morphisme de T_1 dans J en vertu de la construction de la jacobienne. Il résulte de là que l'application $t \to f(t) - f(t_0) + \mathfrak{k}_0$, qui coincide au voisinage de chaque point avec un morphisme d'un voisinage de ce point dans J, est un morphisme de T dans J; prenant en particulier $\mathfrak{k}_0 = 0$, on voit que $t \to f(t) - f(t_0)$ est un morphisme de T dans J.

Supposons réciproquement que $t \to f(t) - f(t_0)$ soit un morphisme de T dans J. Pour montrer que f est algébrique, il suffira évidemment de montrer que l'application identique i de J sur $\mathfrak{G}_0(C)$ est une famille algébrique de classes de diviseurs de C. Si $\mathfrak{c} \in \mathfrak{G}_g(C)$, la restriction $i_\mathfrak{c}$ de i à la sous-variété $V(\mathfrak{c})$ de J définie ci-dessus est une famille algébrique, car on l'obtient en composant l'isomorphisme $\theta_\mathfrak{c}^{-1}$ de $V(\mathfrak{c})$ sur Ω avec la famille algébrique $z \to \mathfrak{d}_g(z) - \mathfrak{c}$. Tenant compte de ce qui a été dit dans la première partie de la démonstration, on voit que, pour tout $\mathfrak{k}_0 \in \mathfrak{G}_0(C)$, l'application $\mathfrak{k} \to \mathfrak{k} + \mathfrak{k}_0$ $(\mathfrak{k} \in V(\mathfrak{c}))$ est un morphisme de $V(\mathfrak{c})$ dans J. Ceci étant vrai pour tout $\mathfrak{c}$, on voit que l'application $\mathfrak{k} \to \mathfrak{k} + \mathfrak{k}_0$ est un morphisme de J dans J ; c'est même un automorphisme de J, l'application $\mathfrak{k} \to \mathfrak{k} - \mathfrak{k}_0$ étant également un morphisme. La variété J, qui admet un groupe transitif d'automorphismes, est donc une variété non singulière. Faisant usage de la Propostitition 3, I, § IV, on déduit alors du fait que les $i_\mathfrak{c}$ sont des familles algébriques de classes de diviseurs qu'il en est de même de f, ce qui démontre le Théorème 1.

Corollaire 1. *La variété J, munie de la structure de groupe de $\mathfrak{G}_0(C)$, est une variété abélienne.*

L'application $(\mathfrak{k}, \mathfrak{k}') \to \mathfrak{k} - \mathfrak{k}'$ de $J \times J$ dans J est une famille algébrique de classes de diviseurs de C, puisque $\mathfrak{k} \to \mathfrak{k}$ en est une. C'est donc un morphisme de $J \times J$ dans J, ce qui montre que J est un groupe algébrique. Soit z_1 un point de S^g; l'application $z \to \mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)$ $(z \in S^g)$ est une

famille algébrique de classes de diviseurs; c'est donc un morphisme de S^g dans J. On sait par ailleurs que ce morphisme est surjectif. Comme S^g est complète, il en est de même de J.

COROLLAIRE 2. *Pour tout $x \in C$, soit $\mathfrak{x}(x)$ la classe du diviseur $1 \cdot x$; si $x_0 \in C$, l'application $x \to \mathfrak{x}(x) - \mathfrak{x}(x_0)$ est un morphisme de C dans J.*

Cela résulte immédiatement du Théorème 1. Une application de C dans J définie de cette manière est appelée *canonique.*

II. Familles de classes de diviseurs parametrées par une courbe. Nous désignerons par C une courbe complète et normale. Pour tout $r > 0$, nous désignerons par S^r la puissance symétrique r-ième de C, par d_r la famille canonique de diviseurs de C paramétrée par S^r, et, pour $z \in S^r$, par $\mathfrak{d}^r(z)$ la classe de $d^r(z)$.

PROPOSITION 1. *Soit $\mathfrak{a}$ une classe de diviseurs de C. Si l'ensemble E des points $z \in S^r$ tels que $\mathfrak{d}_r(z) = \mathfrak{a}$ n'est pas vide, c'est une sous-variété fermée de S^r isomorphe à un espace projectif.*

Supposons $E \neq \emptyset$, et soit z_1 un point de cet ensemble; soit V l'espace vectoriel des fonctions multiples de $-d_r(z_1)$; désignons par $\zeta \to \operatorname{div}\zeta$ le système linéaire de diviseurs de C paramétré par l'espace projectif $\mathfrak{P}(V)$ associé à V. Si $\zeta \in \mathfrak{P}(V)$, il y a un point $f(\zeta)$et un seul de S^r tel que $d_r(f(\zeta)) = \operatorname{div}\zeta + d_r(z_1)$; il est clair que f est une bijection de $\mathfrak{P}(V)$ sur E. Nous allons montrer que c'est un morphisme de $\mathfrak{P}(V)$ dans S^r. Dans le cas où $\mathfrak{a}$ contient un diviseur qui est somme de r points distincts de C, cela résulte de la Proposition 2, I, § III et du fait que d_r est infinitésimalement injective en tout point z tel que $d_r(z)$ soit somme de r points distincts. Pour établir notre assertion dans le cas général, nous observerons qu'il existe un diviseur entier b que la classe de $d_r(z_1) + b$ contienne un diviseur qui soit la somme de $r + r'$ points distincts (où r' est le degré de b). En effet, soit a un diviseur de degré $r + g$ qui soit somme de $r + g$ points distincts; l'espace des fonctions qui sont multiples de $-a$ est de dimension $\geqq r + 1$, d'où on déduit qu'il existe une fonction $\neq 0$ multiple de $-a + d_r(z_1)$; si $-a + d_r(z_1) + b$ est le diviseur de cette fonction, b possède la propriété requise (on peut donc prendre $r' = g$; nous supposons qu'il en est ainsi). Soit Σ l'ensemble des points $t \in S^{r+g}$ tels que $d_{r+g}(t)$ soit multiple de b; il y a une application bijective σ de Σ sur S^r telle que $d_r(\sigma(t)) = d_{r+g}(t) - b \quad (t \in \Sigma)$. Montrons que Σ est une sous-variété fermée de S^{r+g} et que σ est un morphisme. L'ensemble Σ est fermé en vertu du Corollaire 3 au Théorème 1, § I; il y au moins une composante

irréductible Σ' de Σ telle que $\sigma(\Sigma')$ soit dense dans S^r. L'application $t \rightarrow d_{r+g}(t) - b$ induit une famille algébrique de diviseurs de C paramétrée par Σ'; par ailleurs, les points z tels que $d_r(z)$ soit somme de r points distincts forment une partie ouverte non vide de S^r, de sorte que $\sigma(\Sigma')$ rencontre cet ensemble. Faisant usage du résultat cité plus haut, on voit que la restriction de σ à Σ' est un morphisme de Σ' dans S^r. Comme Σ' est fermé dans S^{r+g}, c'est une variété complète; $\sigma(\Sigma')$ est donc fermé, d'où $\sigma(\Sigma') = S^r$ et $\Sigma' = \Sigma$ puisque σ est injectif. Ceci étant, soit V' l'espace des fonctions qui sont multiples de $-d_r(z_1) - b$; soit $\zeta' \rightarrow \operatorname{div} \zeta'$ le système linéaire paramétrée par $\mathfrak{P}(V')$. Il résulte de ce que nous avons dit que l'application f' de $\mathfrak{P}(V')$ dans S^{r+g} définie par la condition que $d_{r+g}(f'(\zeta)) = \operatorname{div} \zeta + d_r(z_1) + b$ $(\zeta \in \mathfrak{P}(V'))$ est un morphisme. Par ailleurs, on a $V \subset V'$, de sorte que $\mathfrak{P}(V)$ est une sous-variété de $\mathfrak{P}(V')$. Il est clair que $f'(\mathfrak{P}(V)) \subset \Sigma$ et que $f(\zeta) = \sigma(f'(\zeta))$ si $\zeta \in \mathfrak{P}(V)$; f est donc bien un morphisme.

Il résulte de là que E est une sous-variété fermée de S^r. Soit g l'application réciproque de f; montrons que g est un morphisme de E dans $\mathfrak{P}(V)$. Si $z \in E$, $g(z)$ est défini par la condition que $\operatorname{div} g(z) = d_r(z) - d_r(z_1)$; comme la famille $\zeta \rightarrow \operatorname{div} \zeta$ est injective et infinitésimalement injective, le fait que g soit un morphisme résulte de la Proposition 2, I, § III. L'application f est donc un isomorphisme de $\mathfrak{P}(V)$ sur E.

PROPOSITION 2. *Soit W une sous-variété ouverte de S^r, et soit f une famille algébrique de classes de diviseurs d'une variété normale U paramétrée par W. Si z, z' sont des points de W tels que $\mathfrak{d}_r(z) = \mathfrak{d}_r(z')$, on a $f(z) = f(z')$.*

Il résulte immédiatement de la Proposition 1 qu'il y a une sous-variété E de S^r isomorphe à la droite projective passant par z et z' (sauf si $z = z'$, auquel cas le résultat est évident). Soit E_0 l'ensemble $E \cap W$; E_0 est isomorphe à une sous-variété ouverte de la droite projective, et la restriction f_0 de f à E_0 est une famille algébrique de classes de diviseurs de U paramétrée par E_0. Or on a $\mathfrak{N}(E_0, U) = \{0\}$ (Proposition 5, I, § IV); f_0 est donc constante, ce qui démontre la Proposition 2, puisque z et z' appartiennent à E_0.

PROPOSITION 3. *Soient J la jacobienne de C et χ une application canonique de C dans J. Soient C_1 une sous-variété ouverte de C et f une famille algébrique de classes de diviseurs d'une variété normale et semi-complète U paramétrée par C_1. Il y a alors une famille algébrique g de classes de diviseurs de U paramétrée par J telle que $g(\chi(x)) = f(x)$ pour tout $x \in C_1$; si $\mathfrak{k}_0$ est un point de J, l'application $\mathfrak{k} \rightarrow g(\mathfrak{k}) - g(\mathfrak{k}_0)$ est un homomorphisme du groupe J dans $\mathfrak{G}(U)$.*

Montrons que, si C_2 est une sous-variété ouverte de C_1 et si la proposition est vraie pour la restriction f_2 de f à C_2, elle est vraie pour f. Soit g une application algébrique de J dans $\mathfrak{G}(U)$ telle que $g(\chi(x)) = f_2(x)$ pour $x \in C_2$; l'application $x \to g(\chi(x))$ $(x \in C_1)$ est une famille algébrique de classes de diviseurs de U paramétrée par C_1 et qui coincide avec f sur C_2; cette famille est donc identique à f (Théorème 2, I, § IV). On peut donc supposer sans restriction de généralité qu'il existe une famille algébrique $\bar{f}$ de diviseurs de U paramétrée par C_1 telle que $f(x) = \mathrm{Cl.}\bar{f}(x)$ pour $x \in C_1$. Soit alors r un entier $\geqq 0$; soit W_r l'ensemble des points z de S^r tels que $d_r(z)$ soit la somme de r points de C_1; c'est une sous-variété ouverte de S^r qui s'identifie à la puissance symétrique r-ième de C_1. Il existe donc une famille algébrique $\bar{h}_r$ de diviseurs de U paramétrée par W_r telle que $\bar{h}_r(z) = \bar{f}(x_1) + \cdots + \bar{f}(x_r)$ si $d_r(z) = x_1 + \cdots + x_r$, avec $x_i \in C_1$ $(1 \leqq i \leqq r)$. Si $z \in W_r$, soit $h_r(z)$ la classe de $\bar{h}_r(z)$; il résulte de la Proposition 2 que la condition $d_r(z) \sim d_r(z')$ $(z, z' \in C_1)$ entraîne $h_r(z) = h_r(z')$.

Soit Ω_1 l'intersection de W_g avec l'ensemble des points $z \in S^g$ tels que $d_g(z)$ soit non spécial. Soit z_1 un point quelconque de W_g; rappelons que l'application $z \to \mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)$ $(z \in \Omega_1)$ est un isomorphisme de la sous-variété ouverte Ω_1 de S^g sur une sous-variété ouverte $V(z_1)$ de J. Il y a donc une application algébrique g'_{z_1} *de* $V(z_1)$ dans $\mathfrak{G}(U)$ telle que

$$g'_{z_1}(\mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)) = h_g(z) - h_g(z_1)$$

pour tout $z \in \Omega_1$. Montrons que, si z_1, z_1' sont des points de W_g, g'_{z_1} coincide avec $g'_{z_1'}$ dans $V(z_1) \cap V(z_1')$. Soient z et z' des points de Ω tels que $\mathfrak{d}_g(z) - \mathfrak{d}_g(z_1) = \mathfrak{d}_g(z') - \mathfrak{d}_g(z_1')$, d'où $\mathfrak{d}_g(z) + \mathfrak{d}_g(z_1') = \mathfrak{d}_g(z') + \mathfrak{d}_g(z_1)$. Il y a des points t, t' de W_{2g} tels que $d_{2g}(t) = d_g(z) + d_g(z_1')$, $d_{2g}(t') = d_g(z') + d_g(z_1)$, et on a $d_{2g}(t) \sim d_{2g}(t')$; il en résulte que $h_{2g}(t) = h_{2g}(t')$. Mais il est clair que $h_{2g}(t) = h_g(z) + h_g(z_1')$, $h_{2g}(t') = h_g(z') + h_g(z_1)$; notre assertion est donc établie.

Tout point de J peut se mettre sous la forme $\mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)$ avec z et z_1 dans Ω_1 (Corollaire 3 au Théorème 1, § I). Il y a donc une application g' de J dans $\mathfrak{G}(U)$ qui prolonge toutes les applications $g_{z_1'}$. Comme J est une variété non singulière, g' est une famille algébrique de classes de diviseurs (Propositions 3, I, § IV). Montrons que c'est un homomorphisme de J dans $\mathfrak{G}(U)$. Soient z_1 un point de W_g et $\mathfrak{k}$, $\mathfrak{k}'$ des éléments de $V(z_1)$; écrivons $\mathfrak{k} = \mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)$, $\mathfrak{k}' = \mathfrak{d}_g(z') - \mathfrak{d}_g(z_1)$, avec $z, z' \in \Omega_1$; on a $\mathfrak{k} - \mathfrak{k}' = \mathfrak{d}_g(z) - \mathfrak{d}_g(z')$, et par suite

$$g'(\mathfrak{k}) = h_g(z) - h_g(z_1), \qquad g'(\mathfrak{k}') = h_g(z') - h_g(z_1),$$
$$g'(\mathfrak{k} - \mathfrak{k}') = h_g(z) - h_g(z') = g'(\mathfrak{k}) - g'(\mathfrak{k}');$$

les applications $(\mathfrak{k}, \mathfrak{k}') \rightarrow g'(\mathfrak{k}-\mathfrak{k}')$, $(\mathfrak{k}, \mathfrak{k}') \rightarrow g'(\mathfrak{k}) - g'(\mathfrak{k}')$ sont des applications algébriques de $J \times J$ dans $\mathfrak{G}(U)$ qui coincident sur l'ensemble ouvert non vide $V(z_1) \times V(z_1)$ et qui sont par suite égales, ce qui démontre que g' est un homomorphisme.

Si $z_1 \in W_g$, l'application $z \rightarrow \mathfrak{d}_g(z) - \mathfrak{d}_g(z_1)$ est un morphisme de S^g tout entier dans J; les applications $z \rightarrow g'(\mathfrak{d}_g(z) - \mathfrak{d}_g(z_1))$ et $z \rightarrow h_g(z) - h_g(z_1)$ $(z \in W_g)$ sont des familles algébriques de classes de diviseurs paramétrées par W_g qui coincident sur Ω_1 et qui sont par suite égales. Si $x \in C$, désignons par $\mathfrak{x}(x)$ la classe de $1 \cdot x$; si $g > 0$, soit m un diviseur entier de degré $g-1$ de support contenu dans C_1; écrivons $m = d_{g-1}(c)$, $c \in S^{g-1}$. Soit x_1 un point de C; prenons pour z_1 le point de W_g tel que $d_g(z_1) = 1 \cdot x_1 + m$; si $x \in C_1$, soit z le point de W_g tel que $d_g(z) = 1 \cdot x + m$. On a

$$\mathfrak{x}(x) - \mathfrak{x}(x_1) = \mathfrak{d}_g(z) - \mathfrak{d}_g(z_1),$$

d'où

$$\begin{aligned} g'(\mathfrak{x}(x) - \mathfrak{x}(x_1)) &= h_g(z) - h_g(z_1) \\ &= (f(x) + h_{g-1}(c)) - (f(x_1) + h_{g-1}(c)) = f(x) - f(x_1). \end{aligned}$$

Par ailleurs, il y a un point $x_0 \in C$ tel que $\chi(x) = \mathfrak{x}(x) - \mathfrak{x}(x_0)$; si nous posons $g(\mathfrak{k}) = g'(\mathfrak{k}) + f(x_1) - g'(\mathfrak{x}(x_1) - \mathfrak{x}(x_0))$, on a $g(\chi(x)) = f(x)$ si $x \in C_1$, ce qui établit l'existence de g dans le cas où C est de genre > 0. Si C est de genre 0, on a $J = \{0\}$ et f est constante, car, si $x, x' \in C_1$, on a $1 \cdot x \sim 1 \cdot x'$, d'où $f(x) = h_1(x) = h_1(x') = f(x')$. De plus, comme g ne diffère de g' que par une constante, on a $g(\mathfrak{k}) - g(\mathfrak{k}_0) = g'(\mathfrak{k})$, de sorte que l'application $\mathfrak{k} \rightarrow g(\mathfrak{k}) - g(\mathfrak{k}_0)$ est un homomorphisme.

Chapitre III.

I. Definition de la notion de variete de Picard. Soient U une variété et G un groupe algébrique. Nous appellerons *homomorphisme algébrique* de G dans $\mathfrak{G}(U)$ un homomorphisme du groupe G dans $\mathfrak{G}(U)$ qui est en même temps une application algébrique de G dans $\mathfrak{G}(U)$.

On dit qu'un couple (P, π) formé d'un groupe algébrique P et d'un homomorphisme algébrique π de P dans $\mathfrak{G}(U)$ est une *variété de Picard* de U si la condition suivante est satisfaite: pour tout groupe algébrique G et tout homomorphisme algébrique γ de G dans $\mathfrak{G}(U)$, il existe un homomorphisme φ et un seul de G dans P tel que $\gamma = \pi \circ \varphi$ [quand nous parlons d'homomorphismes de groupes algébriques, il s'agit bien entendu d'homomorphismes rationnels].

Soient U et U' des variétés qui admettent des variétés de Picard (P,π) et (P',π'), et soit f un morphisme de U' dans U. Alors $x\to f^*(\pi(x))$ est un homomorphisme algébrique de P dans $\mathfrak{G}(U')$; il existe donc un homomorphisme φ_f et un seul de P dans P' tel que $f^*(\pi(x))=\pi'(\varphi_f(x))$ pour tout $x\in P$. Si U'' est une troisiéme variété qui admet une variété de Picard (P'',π''), et si f'' est un morphisme de U'' dans U', il est clair que l'on a $\varphi_{f\circ f'}=\varphi_{f'}\circ\varphi_f$; il en résulte en particulier que, si (P,π) et (P',π') sont des variétés de Picard d'une même variété U, il y a un isomorphisme φ et un seul de P sur P' tel que $\pi=\pi'\circ\varphi$.

Proposition 1. *Si une variété normale et semi-complète U admet une variété de Picard (P,π), l'homomorphisme π est injectif.*

Soit en effet N son noyau. Comme π est une application algébrique, N est un sous-groupe fermé de P (Théorème 2, I, § IV) ; soit ω l'application canonique de P sur P/N. Il y a alors un homomorphisme π^* du groupe P/N dans $\mathfrak{G}(U)$ tel que $\pi^*\circ\omega=\pi$, et π^* est algébrique (Proposition 2, I, § V). Il y a donc un homomorphisme φ de P/N dans P tel que $\pi^*=\pi\circ\varphi$, d'où $\pi=\pi\circ\varphi\circ\omega$. It en résulte que $\varphi\circ\omega$ est l'automorphisme identique de P, de sorte que N se réduit à son élément neutre.

Corollaire. *Une variété de Picard est toujours un groupe commutatif.*

Proposition 2. *Soit U une variété normale et semi-complète. Pour que U admette une variété de Picard, il faut et suffit que la condition suivante soit satisfaite: si (G_n,γ_n) est une suite de couples composés chacun d'un groupe algébrique G_n et d'un homomorphisme injectif algébrique $\gamma_n\colon G_n\to\mathfrak{G}(U)$, et si ω_n est, pour tout n, un homomorphisme de G_n dans G_{n+1} tel que $\gamma_{n+1}\circ\omega_n=\gamma_n$, alors il existe un entier n_0 tel que, pour tout $n\geqq n_0$, γ_n soit un isomorphisme.*

Supposons d'abord que U admette une variété de Picard (G,γ). Il existe alors pour tout n un homomorphisme θ_n de G_n dans G tel que $\gamma_n=\gamma\circ\theta_n$; comme γ_n est injectif, il en est de même de θ_n. On a $\gamma\circ\theta_{n+1}\circ\omega_n=\gamma_{n+1}\circ\omega_n=\gamma_n$; comme θ_n est uniquement determiné par la condition que nous lui avons imposée, on a $\theta_{n+1}\circ\omega_n=\theta_n$. Les sous-groupes fermés $\theta_n(G_n)$ de G forment donc une suite croissante, ce qui montre qu'ils sont tous égaux à partir d'un certain rang n_1; soit G' leur valeur commune. Pour $n\geqq n_1$, ω_n est un morphisme bijectif de G_n sur G_{n+1}; soit d_n son degré. Considérant θ_n, pour $n\geqq n_1$, comme un morphisme bijectif de G_n sur G', on voit que θ_{n_1} peut se représenter (si $n>n_1$) comme composé de θ_n et d'un morphisme bijectif $G_{n_1}\to G_n$ de degré $d_{n_1}\cdots d_{n-1}$ (à savoir le composé de $\omega_{n-1},\cdots,\omega_{n_1}$).

On en conclut que $d_{n_1} \cdots d_{n-1}$ est au plus égal au degré de θ_{n_1}, donc que $d_n = 1$ pour tout n assez grand, soit pour $n \geqq n_0$. Si $n \geqq n_0$, ω_n est un morphisme bijectif et birationnel de G_n sur G_{n+1}; G_{n+1} étant une variété normale, il en résulte que ω_n est un isomorphisme.

Supposons réciproquement la condition satisfaite. Un raisonnement classique montre alors qu'il existe un groupe algébrique P et un homomorphisme algébrique injectif π de P dans $\mathfrak{G}(U)$ qui possèdent la propriété suivante: pour tout couple (G', γ') formé d'un groupe algébrique G' et d'un homomorphisme algébrique injectif γ' de G' dans $\mathfrak{G}(U)$, tout homomorphisme ω de P dans G' tel que $\pi = \gamma' \circ \omega$ est un isomorphisme. On va montrer que (P, π) est une variété de Picard de U. Soient G un groupe algébrique et γ un homomorphisme algébrique de G dans $\mathfrak{G}(U)$. L'application $(x, s) \rightarrow \pi(x) + \gamma(s)$ $(x \in P, s \in G)$ est un homomorphisme algébrique de $P \times G$ dans $\mathfrak{G}(U)$; soit N son noyau; soient G' le groupe $(P \times G)/N$ et ζ l'application canonique de $P \times G$ sur G. On sait que N est un sous-groupe fermé et que l'application γ' de G' dans $\mathfrak{G}(U)$ définie par la condition que $\gamma'(\zeta(x, s)) = \pi(x) + \gamma(s)$ $((x, s) \in P \times G)$ est un homomorphisme algébrique (Proposition 2, I, § V); cet homomorphisme est injectif. Soit e_G l'élément neutre de G; soit ω l'application $x \rightarrow \zeta(x, e_G)$ de P dans G'. Il est clair que ω est un homomorphisme et que $\pi = \gamma' \circ \omega$; ω est donc un isomorphisme. Soit e_P l'élément neutre de P; en composant avec ω^{-1} l'homomorphisme $s \rightarrow \omega(e_P, s)$ de G dans G', on obtient un homomorphisme φ de G dans P; il est clair que $\pi = \gamma \circ \varphi$. Comme π est injectif, il n'y a qu'un seul homomorphisme φ de G dans P tel que $\pi = \gamma \circ \varphi$; (P, π) est donc bien une variété de Picard de U.

THÉORÈME 1. *Soit U une variété normale semi-complète qui admet une variété de Picard (P, π). Si f est une famille algébrique de classes de diviseurs de U paramétrée par une variété normale T, et si t_0 est un point de T, il existe un morphisme g et un seul de T dans P tel que l'on ait $f(t) = \pi(g(t)) + f(t_0)$ pour tout $t \in T$; de plus, P est une variété complète.*

Nous considérerons d'abord le cas où T est une courbe. Comme T est normale, il est bien connu que T est isomorphe à une sous-variété ouverte d'une courbe normale et complète C; soit J la jacobienne de C. Faisant usage de la Proposition 3, II, § II, on voit que, si χ est une application canonique de C dans J, il y a un homomorphisme algébrique f_1 de J dans $\mathfrak{G}(U)$ tel que l'on ait $f(t) = f_1(\chi(t)) + \mathfrak{c}_0$ pour tout $t \in T$, $\mathfrak{c}_0$ étant un certain point de $\mathfrak{G}(U)$. Il existe un homomorphisme φ de J dans P tell que $f_1 = \pi \circ \varphi$. Posons $g(t) = \varphi(\chi(t)) - \varphi(\chi(t_0))$ $(t \in T)$; il est clair que g est un morphisme de T dans P, et l'on a, si $t \in T$,

$$f(t)-f(t_0)=f_1(\chi(t)-\chi(t_0))=\pi(g(t)).$$

Le morphisme g est caractérisé de manière unique par cette condition puisque π est injectif. Ceci démontre la première assertion dans le cas où T est une courbe. De plus, on notera que $f_1(J)$ est un sous-groupe *abélien* (i. e. complet) de P.

Comme toute suite croissante de sous-groupes connexes complets de P est constante à partir d'un certain rang, il existe un sous-groupe abélien maximal P' de P. Si P'_1 est un sous-groupe abélien quelconque de P, le groupe $P'+P'_1$, image par un homomorphisme du groupe abélien $P'\times P'_1$, est abélien, donc identique à P', ce qui montre que $P'_1\subset P'$. Il résulte de la première partie de la démonstration que, si f est une application algébrique d'une courbe normale T dans $\mathfrak{G}(U)$, on a $f(t)-f(t')\in\pi(P')$ quels que soient t et t' dans T.

Soit maintenant T une variété normale quelconque; soient f une application algébrique de T dans $\mathfrak{G}(U)$ et t_0 un point de T. Montrons que l'on a $f(t)-f(t_0)\in\pi(P')$ pour tout $t\in T$. Soit A l'ensemble des points t tels que $f(t)-f(t_0)\in\pi(P')$; il suffira de montrer que A est dense dans T. En effet, l'ensemble des points $(t,x)\in T\times P'$ tels que $f(t)-f(t_0)=\pi(x)$ est fermé; comme P' est complète, il en résulte que l'ensemble A est fermé. Montrons que toute courbe Γ contenue dans T et passant par t_0 est contenue dans A. Il existe une courbe normale Γ' et un morphisme r de Γ' dans Γ tels que (Γ',r) soit un revêtement de Γ. L'application $t'\to f(r(t'))$ $(t'\in\Gamma')$ est une application algébrique de Γ' dans $\mathfrak{G}(U)$. Par ailleurs, si $t\in\Gamma$, il existe toujours un point $t'\in\Gamma'$ tel que $r(t')=t$; désignant par t_0' un point tel que $r(t_0')=t_0$, on a $f(t)-f(t_0)=f(r(t'))-f(r(t_0'))\in\pi(P')$ en vertu de ce qui a été dit plus haut. Pour montrer que $A=T$, il suffira donc de montrer que la réunion des courbes tracées sur T et passant par t_0 est une partie dense de T. Mais cela résulte du fait que, si E est une partie fermée $\neq T$ de T, il existe au moins une courbe tracée sur T, passant par t_0 et non contenue dans E ([2], chap. III, § III, Proposition 1).

Appliquons ceci au cas où $T=P$, $f=\pi$, t_0 étant l'élément neutre de P; on a $\pi(P)\subset\pi(P')$, d'où $P=P'$ puisque π est injectif. Ceci montre que P est une variété complète.

Revenons à la consideration de la variété T. Soit H l'ensemble des points $(t,x)\in T\times P$ tels que $f(t)-f(t_0)=\pi(x)$. On sait que cet ensemble est fermé. Il résulte de ce qu'on vient de dire que la restriction p à H de la projection $T\times P\to T$ est une application surjective de H sur T. Comme π est injectif, l'application p est en fait bijective. Si H_1 est une composante

irréductible de H telle que $p(H_1)$ soit dense dans T, $p(H_1)$ est égal à T comme il résulte de ce que H_1 est fermé dans $T \times P$ et de ce que P est complète; comme p est bijectif, on a $H_1 = H$; H est donc une sous-variété fermée de $T \times P$ et p est un morphisme bijectif de H sur T. Si nous montrons que p est birationnel, le théorème sera établi; en effet, composant p^{-1} avec la restriction à H de la projection $T \times P \to P$, on obtiendra un morphisme g de T dans P tel que $f(t) - f(t_0) = \pi(g(t))$ pour tout $t \in T$, et g sera uniquement determiné par cette condition puisque π est injectif.

Observons d'abord que, si Δ est une courbe sur H telle que $\Gamma = p(\Delta)$ soit une courbe normale, la restriction p_Δ de p à Δ est un morphisme birationnel de Δ sur Γ. Soit en effet t_1 un point de Γ. Il existe un morphisme g_1 de Γ dans P tel que l'on ait $f(t) - f(t_1) = \pi(g_1(t))$ pour tout $t \in \Gamma$. Comme $f(t_1) - f(t_0) \in \pi(P)$, il en résulte qu'il y a un morphisme g_2 de Γ dans P tel que l'on ait $f(t) - f(t_0) = \pi(g_2(t))$ $(t \in \Gamma)$; Δ n'est alors autre que l'ensemble des points $(t, g_2(t))$ pour $t \in \Gamma$, ce qui montre que p_Δ est un isomorphisme de Δ sur Γ. Nous sommes donc ramenés à établir le lemme suivant:

LEMME 1. *Soit p un morphisme bijectif d'une variété H dans une variété T. Si p est de degré > 1, il existe une courbe Δ sur H telle que $p(\Delta)$ soit normale et que la restriction de p à Δ soit de degré > 1.*

Soit π le cohomomorphisme de p; soient $F(T)$ et $F(H)$ les corps des fonctions numériques sur T et sur H. Désignons par q la caractéristique de K, qui est > 0 en vertu des hypothèses faites. Comme p est radiciel, il y a une fonction $\theta \in F(T)$ qui n'est pas puissance q-ième dans $F(T)$ mais qui est telle que $\pi(\theta)$ soit puissance q-ième dans $F(H)$. On sait qu'il existe alors une dérivation X du corps $F(T)$ telle que $X(\theta) \neq 0$. A cette dérivation est associé un champ de vecteurs tangents à T défini sur une partie ouverte non vide T_1 de T, champ de vecteurs que nous désignerons encore par X. Soient par ailleurs H_0 et T_0 les ensembles de points simples de H et T; $p(H_0)$ contient une partie ouverte non vide de T. Il en résulte qu'il y a un point $t_0 \in p(H_0) \cap T_0 \cap T_1$ tel que θ soit définie en t_0 et que $(X(\theta))(t_0) \neq 0$, ce qui signifie que $< X(t_0), \theta > \neq 0$. Comme t_0 est un point simple de T, il y a une courbe Γ_1 de T, passant par t_0, y admettant un point simple, telle que $X(t_0)$ soit tangent à Γ_1 en t_0. Si θ_1 est l'empreinte de θ sur cette courbe, on a $< X(t_0), \theta_1 > \neq 0$, ce qui montre que θ_1 n'est pas puissance p-ième dans le corps des fonctions numériques sur Γ_1. Comme T est normale en t_0 et comme $p^{-1}(t_0)$ se compose d'un seul point t_0', il y a une courbe Δ_1 passant par t_0' telle que $p(\Delta_1)$ soit dense dans Γ_1 ([2], chap. V, § V, Proposition 2). Soit p_1 la restriction de p à Δ_1. Montrons que $\theta_1 \circ p_1$ est puissance q-ième dans

le corps des fonction numériques sur Δ_1. La fonction $\theta \odot p$ est puissance q-ième d'une fonction θ' sur H ; de plus, elle est définie au point t_0' en lequel H est normale puisque $t_0' \in H_0$; θ' est donc définie en t_0' et admet une empreinte θ'_1 sur Δ_1 ; il est alors clair que $\theta_1 \odot p_1 = \theta'_1{}^q$. Ceci montre que p_1 est de degré > 1. L'ensemble $p_1(\Delta_1)$ contient une sous-variété ouverte normale Γ de Γ_1 ; il suffit alors de prendre $\Delta = p^{-1}(\Gamma)$.

Remarque. Soit U une variété normale et complète sur laquelle on fait l'hypothèse suivante : il existe un groupe algébrique *complet* P et un homomorphisme algébrique injectif π de P dans $\mathfrak{G}(U)$ tels que, pour tout groupe algébrique *complet* G et tout homomorphisme algébrique γ de G dans $\mathfrak{G}(U)$, il existe un homomorphisme φ et un seul de G dans P tel que $\gamma = \pi \circ \varphi$. Alors (P, π) est variété de Picard de U. Reprenons en effet la démonstration précédente. La jacobienne d'une courbe étant une variété complète, la première partie de la démonstration établit encore que, si T est une courbe normale et f une application algébrique de T dans $\mathfrak{G}(U)$, et si $t_0 \in T$, il existe un morphisme g et un seul de T dans P tel que $f(t) - f(t_0) = \pi(g(t))$ pour tout $t \in T$. Le reste de la démonstration se transporte alors sans modification, et établit que, si T est une variété normale quelconque et f une application algébrique de T dans $\mathfrak{G}(U)$, et si $t_0 \in T$, il existe un morphisme g et un seul de T dans P tel que $f(t) - f(t_0) = \pi(g(t))$ $(t \in T)$. Appliquons ceci au cas où T est un groupe algébrique et g un homomorphisme algébrique, t_0 étant l'élément neutre du groupe, d'où $f(t_0) = 0$; comme π est un homomorphisme injectif, il en résultera que g est un homomorphisme ; (P, π) est donc bien une variété de Picard de U. Ceci étant, on voit que, si U est normale et semi-complète, il suffit pour que U admette une variété de Picard, que la condition énoncée dans la Proposition 2 soit satisfaite quand on y suppose de plus que les G_n sont des groupes complets ; il suffit en effet de reprendre la démonstration de la Proposition 2 en n'y considérant que des groupes complets : on montre ainsi qu'il existe un groupe complet P et un homomorphisme algébrique injectif π de P dans $\mathfrak{G}(U)$ qui possèdent la propriété énoncée au début de cette remarque.

Proposition 3. *Soient U et U' des variétés normales et complètes. Supposons que U' admette une variété de Picard (P', π') et qu'il existe un morphisme surjectif f de U' dans U. Alors U admet une variété de Picard.*

Soient G un groupe algébrique complet et γ un homomorphisme algébrique injectif de G dans $\mathfrak{G}(U)$. Alors l'application $\gamma' : s \to f^*(\gamma(s))$ est un homomorphisme algébrique de G dans $\mathfrak{G}(U')$. Nous allons montrer qu'il est de noyau fini. Soit N la composante algébrique de l'élément neutre dans son

noyau; N est une variété complète. Il en résulte que l'homomorphisme $f^*: \mathfrak{N}(U,N) \to \mathfrak{N}(U',N)$ est injectif (Proposition 4, I, § IV). En vertu des isomorphismes canoniques $\mathfrak{N}(U,N) \cong \mathfrak{N}(N,U)$, $\mathfrak{N}(U',N) \cong \mathfrak{N}(N,U')$, l'homomorphismes $f^*: \mathfrak{N}(N,U) \to \mathfrak{N}(N,U')$ est injectif. Soit γ_0 la restriction de γ à N; cette application de N dans $\mathfrak{G}(U)$ est définie par un élément $m \in \mathfrak{M}(N,U)$ dont l'image par l'homomorphisme $f: \mathfrak{M}(N,U) \to \mathfrak{M}(N,U')$ est nulle. Il résulte alors de ce que nous venons de dire que l'image de m dans $\mathfrak{N}(N,U)$ est nulle, c'est-à-dire que γ_0 est constante. Comme c'est un homomorphisme, c'est l'application nulle; comme γ est injectif, N se réduit à son élément neutre. Ceci étant, soit (G_n, γ_n) une suite de couples formés chacun d'un groupe algébrique complet G_n et d'un homomorphisme algébrique injectif γ_n de G_n dans $\mathfrak{G}(U)$; soit de plus, pour tout n, ω_n un homomorphisme de G_n dans G_{n+1} tel que $\gamma_n = \gamma_{n+1} \circ \omega_n$. Soit γ_n' l'application $s \to f^*(\gamma_n(s))$ de G_n dans $\mathfrak{G}(U')$; il existe donc un homomorphisme algébrique θ_n et un seul de G_n dans P' tel que $\gamma_n' = \pi' \circ \theta_n$. Il est clair que l'on a $\theta_n = \theta_{n+1} \circ \omega_n$. Soit N_n le noyau de θ_n; on a donc $\omega_n(N_n) \subset N_{n+1}$, et ω_n définit par passage aux quotients un homomorphisme ω_n' du groupe $G_n' = G_n/N_n$ dans G_{n+1}'. Par ailleurs θ_n définit par passage aux quotients un homomorphisme injectif θ_n' de G_n' dans P', et on a $\theta_n' = \theta_{n+1}' \circ \omega_n'$. Appliquant alors la Proposition 2 aux homomorphismes algébriques $\pi' \circ \theta_n'$ des G_n' dans $\mathfrak{G}(U')$, on voit qu'il existe un n_1 tel que ω_n' soit un isomorphisme pour tout $n \geqq n_1$. Si donc $n \geqq n_1$, on a $\dim G_{n+1} = \dim G_{n+1}' = \dim G_n' = \dim G_n$, et, comme ω_n est évidemment injectif (les γ_n l'étant), il en résulte que ω_n est aussi surjectif. L'homomorphisme obtenu en composant ω_n avec l'application canonique $G_{n+1} \to G_{n+1}'$ s'obtient aussi en composante l'application canonique $G_n \to G_n'$ avec l'isomorphisme ω_n'; il est donc séparable, d'où il résulte que ω_n est luimême séparable. Etant bijectif, il est birationnel; comme G_{n+1} est normale, ω_n est un isomorphisme. Tenant compte de la remarque qui suit la démonstration du Théorème 1, on voit que U admet une variété de Picard.

II. Variete de Picard d'un produit. Soit U une variété normale et semicomplète qui admet une variété de Picard (P, π). Soit T une variété normale; nous identifierons les applications algébriques de T dans $\mathfrak{G}(U)$ aux éléments du groupe $\mathfrak{M}(T,U)$. Si t_0 est un point de T, nous désigerons par $\mathfrak{M}(t_0; T, U)$ le groupe des applications algébriques de T dans $\mathfrak{G}(U)$ qui appliquent t_0 sur 0. Ce groupe est canoniquement isomorphe à $\mathfrak{N}(T,U)$. En effet, le noyau de l'application canonique $\mathfrak{M}(T,U) \to \mathfrak{N}(T,U)$ se compose des applications constantes de T dans $\mathfrak{G}(U)$ et n'a par suite que 0 en commun avec $\mathfrak{M}(t_0; T, U)$. Par ailleurs si $f \in \mathfrak{M}(T,U)$, la formule $f(t) = (f(t) - f(t_0)) + f(t_0)$ montre

que l'image de f dans $\mathfrak{N}(T, U)$ est aussi l'image d'un élément de $\mathfrak{M}(t_0; T, U)$. Il résulte du Théorème 1, § I que l'application $g \to \pi \circ g$ est un isomorphisme du groupe $\mathrm{Mor}(t_0; T, P)$ des morphismes de T dans P qui appliquenet t_0 sur 0 sur le groupe $\mathfrak{M}(t_0; T, U)$. Les trois groupes

$$\mathfrak{N}(T, U), \qquad \mathfrak{M}(t_0; T, U); \qquad \mathrm{Mor}(t_0; T, P)$$

sont donc canoniquement isomorphes les uns aux autres, Par ailleurs, si on se donne des morphismes f d'une variété normale T' dans T et g d'une variété semi-complète et normale U' dans U, ainsi qu'un point $t_0' \in T'$ tel que $f(t_0') = t_0$, f et g définissent un homomorphisme $r_1 : \mathfrak{N}(T, U) \to \mathfrak{N}(T', U')$; de plus, si $\varphi \in \mathfrak{M}(t_0; T, U)$, l'application $t' \to g^*(\varphi(f(t')))$ appartient à $\mathfrak{M}(t_0'; T', U')$, ce qui définit un homomorphisme $r_2 : \mathfrak{M}(t_0; T, U) \to \mathfrak{M}(t_0'; T', U')$. Enfin, si U' admet une variété de Picard (P', π'), g définit un homomorphisme $\bar{g}$: $P \to P'$, et l'application $\varphi \to \bar{g} \circ \varphi \circ f$ est un homomorphisme r_3 de $\mathrm{Mor}(t_0; T, P)$ dans $\mathrm{Mor}(t_0'; T', P')$. On vérifie immédiatement que le diagramme

$$\begin{tikzcd}
\mathfrak{N}(T, U) \arrow[r] \arrow[d, "r_1"'] & \mathfrak{M}(t_0; T, U) \arrow[r] \arrow[d, "r_2"'] & \mathrm{Mor}(t_0; T, P) \arrow[d, "r_3"'] \\
\mathfrak{N}(T', U') \arrow[r] & \mathfrak{M}(t_0'; T', U') \arrow[r] & \mathrm{Mor}(t_0'; T', P'),
\end{tikzcd}$$

où les lignes horizontales sont données par les homomorphismes canoniques mentionnés ci-dessus, est commutatif.

PROPOSITION 1. *Soient U, V, W des variétés normales dont l'une au moins est semi-complète et admet une variété de Picard; soient p, q, r les projections de $U \times V \times W$ sur $U \times V$, $V \times W$ et $W \times U$ respectivement. Le groupe $\mathfrak{G}(U \times V \times W)$ est alors la somme des images par p^*, q^*, r^* des groupes $\mathfrak{G}(U \times V)$, $\mathfrak{G}(V \times W)$ et $\mathfrak{G}(W \times U)$.*

Supposons que W soit semi-complète et admette une variété de Picard (P, π). Si r_W est la projection $U \times V \times W \to W$, $\mathfrak{N}(U \times V, W)$ est

$$\mathfrak{G}(U \times V \times W) / (p^*(\mathfrak{G}(U \times V)) + r_W^*(\mathfrak{G}(W))).$$

Choisissons par ailleurs un point $x_0 \in U$ et un point $y_0 \in V$; $\mathfrak{N}(U \times V, W)$ est alors isomorphe à $\mathrm{Mor}((x_0, y_0); U \times V, P)$. Or, P étant une variété abélienne, si φ est une fonction sur $U \times V$ à valeurs dans P, il existe des fonctions φ_U sur U et φ_V sur V à valeurs dans P telles que l'on ait $\varphi(x, y) = \varphi_U(x) + \varphi_V(y)$ pour tout point simple $(x, y) \in U \times V$. Si φ est un morphisme, φ_U et φ_V sont des morphismes; en effet, y_1 étant un point simple de V, φ_U coincide évidemment avec le morphisme $x \to \varphi(x, y_1) - \varphi_V(y_1)$, et on voit de même que φ_V est un morphisme. De plus, si on suppose que

$\varphi(x_0, y_0) = 0$, on peut supposer que $\varphi_U(x_0) = \varphi_V(y_0) = 0$, et φ_U et φ_V sont alors uniquement determinés. Soient s_U et s_V les projections de $U \times V$ sur U et sur V; il résulte de ce qu'on vient de dire que l'application $(\varphi_U, \varphi_V) \to \varphi_U \circ s_U + \varphi_V \circ s_V$ est un isomorphisme du groupe

$$\mathrm{Mor}(x_0; U, P) \times \mathrm{Mor}(y_0; V, P) \text{ sur } \mathrm{Mor}((x_0, y_0); U \times V, P).$$

On en conclut que $\mathfrak{N}(U \times V, W)$ est somme directe des groups $s_U^*(\mathfrak{N}(U, W))$ et $s_V^*(\mathfrak{N}(V, W))$, et par suite que $\mathfrak{G}(U \times V \times W)$ est somme des groupes $p^*(\mathfrak{G}(U \times V))$, $q^*(\mathfrak{G}(V \times W))$, $r^*(\mathfrak{G}(W \times U))$ et $r_W^*(\mathfrak{G}(W))$; le dernier de ces groupes étant contenu dans le deuxième (et dans le troisième), la Proposition 1 est établie.

THÉORÈME 2. *Soient U_1 et U_2 des variétés normales et semi-complètes qui admettent des variétés de Picard (P_1, π_1) et (P_2, π_2). Soit q_i la projection de $U_1 \times U_2$ sur U_i $(i = 1, 2)$; soit π l'homomorphisme algébrique*

$$(x_1, x_2) \to q_1^*(\pi_1(x_1)) + q_2^*(\pi_2(x_2))$$

de $P_1 \times P_2$ dans $\mathfrak{G}(U_1 \times U_2)$. Alors $(P_1 \times P_2, \pi)$ est une variété de Picard de $U_1 \times U_2$.

Soit G un groupe algébrique. Il résulte immédiatement de la proposition 1 que $\mathfrak{N}(G, U_1 \times U_2)$ est somme des groupes $q_i^*(\mathfrak{N}(G, U_i))$ $(i = 1, 2)$. Soit γ un homomorphisme algébrique de G dans $\mathfrak{G}(U_1 \times U_2)$; il résulte de ce qu'on vient de dire qu'il existe des applications algébriques γ_1 et γ_2 de G dans $\mathfrak{G}(U_1)$ et $\mathfrak{G}(U_2)$ et un élément $\mathfrak{c} \in \mathfrak{G}(U_1 \times U_2)$ tels que

$$\gamma(s) = q_1^*(\gamma_1(s)) + q_2^*(\gamma_2(s)) + \mathfrak{c}$$

pour tout $s \in G$. Soit e l'élément neutre de G; on peut évidemment supposer que $\gamma_1(e) = \gamma_2(e) = 0$, et on a alors $\mathfrak{c} = 0$. Soit x_i un point de U_i $(i = 1, 2)$; soit j_1 (resp. j_2) l'application $y_1 \to (y_1, x_2)$ (resp. $y_2 \to (x_1, y_2)$) de U_1 (resp. U_2) dans $U_1 \times U_2$. Alors $q_2 \circ j_1$ est l'application constante de valeur x_2 de U_1 dans U_2. Il en résulte que

$$j_1^*(q_1^*(\gamma_1(s))) = \gamma_1(s), \qquad j_1^*(q_2^*(\gamma_2(s))) = 0,$$

et par suite que $\gamma_1(s) = j_1^*(\gamma(s))$, ce qui montre que γ_1 est un homomorphisme; on verrait de même que γ_2 est un homomorphisme. Il existe donc un homomorphisme $g_i \colon G \to P_i$ tel que $\gamma_i = \pi_i \circ g_i$ $(i = 1, 2)$. L'application $g \colon s \to (g_1(s), g_2(s))$ est un homomorphisme de G dans $P_1 \times P_2$, et il est clair que $\gamma = \pi \circ g$. Pour montrer que g est uniquement determiné par cette condition, il suffit d'établir que π est injectif; or cela résulte immédiatement des formules $j_i^*(\pi(x_1, x_2)) = \pi_i(x_i)$ $(i = 1, 2)$. Le théorème 2 est donc établi.

Soient C une courbe normale et complète et J sa jacobienne. Désignons par ι l'application identique de J dans $\mathfrak{G}(C)$; alors (J, ι) est une variété de Picard de C. En effet, il est clair que ι est un homomorphisme algébrique de J dans $\mathfrak{G}(C)$. Soient maintenant G un groupe algébrique, e son élément neutre et γ un homomorphisme algébrique de G dans $\mathfrak{G}(C)$. Il résulte du théorème 1, § I que γ est un morphisme de G dans J; comme ι est injectif, γ est un homomorphisme de G dans J et est le seul homomorphisme g de G dans J tel que $\iota \circ g = \gamma$, ce qui établit notre assertion.

Il résulte alors du théorème 2 que tout produit de courbes normales et complètes admet une variété de Picard. Soit maintenant U une variété abélienne; elle peut être munie d'une structure de groupe commutatif, dont nous désignerons l'élément neutre par x_0. Montrons que, si $r = \dim U$, il existe des courbes complètes $\Gamma_1, \cdots, \Gamma_r$ tracées sur U telles qu'il existe un morphisme surjectif de $\Gamma_1 \times \cdots \times \Gamma_r$ sur U. Supposons déjà construites des courbes complètes Γ_i pour $i < k$, tracées sur U et passant par x_0, telles que l'application $(x_i)_{i<k} \rightarrow \sum_{i<k} x_i$ applique $\prod_{i<k} \Gamma_i$ sur une sous-variété U_{k-1} de dimension $k-1$ de U (k étant un entier entre 1 et r). Comme $k-1 < r$, il y a une courbe Γ_k sur U, passant par x_0 mais non contenue dans U_{k-1}; on peut supposer Γ_k fermée dans U, et Γ_k est alors complète. L'image de $\prod_{i \leqq k} \Gamma_i$ par l'application $(x_i)_{i \leqq k} \rightarrow \sum_{i \leqq k} x_i$ est une sous-variété U_k de U (puisque les Γ_i sont complètes) qui contient U_{k-1} et Γ_k (puisque les Γ_i passent par x_0) et qui est par suite de dimension $\geqq k$; comme $\prod_{i \leqq k} \Gamma_i$ est de dimension k, U_k est de dimension $\leqq k$. Ceci étant, on a $U_r = U$, ce qui démontre notre assertion. Pour tout $i \leqq r$, il existe un morphisme surjectif d'une courbe normale et complète C_i sur Γ_i; il existe donc un morphisme surjectif de la variété $C_1 \times \cdots \times C_r$ sur U; tenant compte de la proposition 3, § I, on en déduit la

PROPOSITION 2. *Toute variété abélienne admet une variété de Picard.*

III. Variete d'Albanese stricte. Soit U une variété. Il est bien connu qu'il existe une variété abélienne A et une fonction dominante f sur U à valeurs dans A qui possèdent la propriété suivante: si g est une fonction quelconque sur U à valeurs dans une variété abélienne B, il existe un morphisme h et un seul de A dans B tel que $g = h \odot f$; de plus, si (A', f') est un autre couple qui possède les mêmes propriétés que (A, f), il existe un isomorphisme j et un seul de A sur A' tel que $f' = j \odot f$; on dit que (A, f) est une *variété d'Albanese* de U. Nous allons voir maintenant qu'il existe un énoncé analogue relatif au cas où on considère des morphismes de U dans

des variétés abéliennes au lieu de fonctions. Soit (A, f) une variété d'Albanese de A; A peut être munie d'une structure de groupe commutatif, dont nous dsignerons l'élément neutre par e. Soit C la correspondance entre U et A associée à f; c'est l'adhérence dans $U \times A$ du graphe de f. Pour tout $x \in U$, soit $E(x)$ l'ensemble des points $a \in A$ tels que $(a, x) \in C$; si f est définie en x (en particulier, si x est simple sur U), $E(x)$ se compose du seul point $f(x)$. Nous désignerons par $E'(x)$ l'ensemble des points de la forme $a' - a$, avec a et a' dans $E(x)$; enfin, nous désignerons par N le plus petit sous-groupe fermé de A contenant les ensembles $E'(x)$ pour tous les $x \in U$. Soit A' la variété abélienne A/N, et soit ω l'application canonique de A sur A'. Posons $f' = \omega \odot f$; montrons que f' est un morphisme de U dans A'. Soit C' la correspondance entre A' et U associée à f'. L'application $\omega_1: (a, x) \rightarrow (\omega(a), x)$ de $A \times U$ dans $A' \times U$ est propre puisque ω est propre (en tant que morphisme d'une variété complète); $\omega_1(C)$ est donc une sous-variété fermée de $A' \times U$. Cette variété est l'adhérence de l'ensemble des points $(f'(x), x)$, x parcourant l'ensemble des points en lesquels f est définie; il en résulte immédiatement que c'est aussi l'adhérence du graphe de f', donc que $C' = \omega_1(C)$. Soit x un point quelconque de U; comme A' est complète, il y a au moins un point $a' \in A'$ tel que $(a', x) \in C'$. Montrons qu'il n'y en a qu'un. Soient a_1' et a_2' des points de A' tels que $(a_i', x) \in C'$ $(i = 1, 2)$. Comme $C' = \omega_1(C)$, il y a des points a_i $(i = 1, 2)$ de A tels que $(a_i, x) \in C$, $a_i' = \omega(a_i)$; mais on a alors $a_2 - a_1 \in E'(x) \subset N$, d'où $a_1' = a_2'$, ce qui établit notre assertion. Ceci étant, il résulte du fait que A' est normale et du théorème principal de Zariski que f' est partout définie, donc que c'est un morphisme. Soit maintenant g un morphisme de U dans une variété abélienne B; soit h le morphisme de A dans B tel que $g = h \odot f$. Soient x un point de U et a_1, a_2 des points de $E(x)$; comme h est définie en ces points, les points $(h(a_1), x)$ et $(h(a_2), x)$ appartiennent à la correspondance entre B et U associée à $h \odot f$ ([2], chap. IV, § I, Proposition 5). Comme $h \odot f = g$ est un morphisme, il en résulte que $h(a_1) = h(a_2) = g(x)$. Or B peut être munie d'une structure de groupe commutatif admettant $h(e)$ comme élément neutre; il est alors bien connu que h est un homomorphisme de A dans B, d'où $h(a_1 - a_2) = 0$. Le noyau de h contient donc tous les ensembles $E'(x)$, ce qui montre qu'il contient N. On en déduit que h se met sous la forme $h' \circ \omega$, où h' est un morphisme de A' dans B; il est clair que l'on a $g = h' \circ f'$. Comme il n'existe qu'un seul morphisme h de A dans B tel que $g = h \circ f$, il n'existe qu'un seul morphisme h' de A' dans B tel que $g = h' \circ f'$. Nous avons donc établi le

THÉORÈME 3. *Soit U une variété. Il existe une variété abélienne A et un morphisme f de U dans A qui possèdent la propriété suivante: si g est*

un morphisme de U dans une variété abélienne B, il existe un morphisme h et un seul de A dans B tel que $g = h \circ f$.

Let notations étant celles du théorème précédent, nous dirons que (A, f) est une *variété d'Albanese stricte de la variété U.*

On notera que, si (A_0, f_0) est une variété d'Albanese de U et (A, f) une variété d'Albanese stricte, A est une variété quotient de A_0; par ailleurs, si U est non singulière, on a $A = A_0$. Le nombre $\nu(U) = \dim A_0 - \dim A$ est donce un indicateur de l'importance des singularités de U. Il n'est pas sans interêt à ce sujet d'observer qu'il existe toujours une variété U' telle que $\nu(U') = 0$ qui admet un morphisme surjectif birationnel sur U: il suffit en effet de prendre pour U' la correspondance entre A_0 et U associée à la fonction f_0. Il y aurait peut être lieu d'examiner si l'opération qui consiste à passer de U à une variété telle que U' ne serait pas utile dans l'étude du problème de la réduction des singularités de U.

IV. Variete de Picard d'une variete normale complete.

THÉORÈME 4. *Toute variété normale semi-complète U admet une variété de Picard.*

Soit (A, f) une variété d'Albanese stricte de U. Comme A est une variété abélienne, elle admet une variété de Picard (P, α). L'application $\pi: z \to f^*(\alpha(z))$ $(z \in P)$ est un homomorphisme algébrique de P dans $\mathfrak{G}(U)$. Nous allons montrer que (P, π) est une variété de Picard de U.

Soit G un groupe algébrique complet. Le morphisme $f: U \to A$ définit un homomorphisme $f^*: \mathfrak{N}(G, A) \to \mathfrak{N}(G, U)$; montrons que set homomorphisme est un isomorphisme. Il suffit de montrer que l'homomorphisme

$$f^*: \mathfrak{N}(A, G) \to \mathfrak{N}(U, G)$$

est un isomorphisme. Nous désignerons par x_0 un point de U; la variété A possède une structure de groupe commutatif admettant $f(x_0) = e$ comme élément neutre. Par ailleurs, G, qui est une variété abélienne, admet une variété de Picard (Q, χ). Il résulte alors de ce qui a été dite au début du § II qu'il existe des isomorphismes canoniques $\mathfrak{N}(A, G) \to \mathrm{Mor}(e; A, Q)$, $\mathfrak{N}(U, G) \to \mathrm{Mor}(x_0; U, Q)$ tels que le diagramme

$$\begin{array}{ccc} \mathfrak{N}(A, G) & \xrightarrow{f^*} & \mathfrak{N}(U, G) \\ \downarrow & & \downarrow \\ \mathrm{Mor}(e; A, Q) & \longrightarrow & \mathrm{Mor}(x_0; U, Q) \end{array}$$

soit commutatif, la deuxième flèche horizontale de ce diagramme étant l'application $h \to h \circ f$ $(h \in \mathrm{Mor}(e; A, Q))$. Or il résulte de la définition des variétés d'Albanese strictes que l'application $h \to h \circ f$ est un isomorphisme de $\mathrm{Mor}(e; A, Q)$ sur $\mathrm{Mor}(x_0; U, Q)$. Il en résulte bien que f^* est un isomorphisme.

Puisque (P, α) est une variété de Picard de A, il y a un isomorphisme canonique de $\mathfrak{N}(G, A)$ sur $\mathrm{Mor}(e; G, P)$, d'où, en tenant compte de l'isomorphisme f^*, un isomorphisme

$$\mathrm{Mor}(e; G, P) \cong \mathfrak{N}(G, U).$$

Cet isomorphisme s'explicite comme suit; $\mathfrak{N}(G, U)$ étant identifié à un groupe quotient de $\mathfrak{M}(G, U)$, donc à un quotient du groupe des applications algébriques de G dans $\mathfrak{G}(U)$, l'élément de $\mathfrak{N}(G, U)$ qui correspond à un élément g de $\mathrm{Mor}(e; G, P)$ est la classe dans $\mathfrak{N}(G, U)$ de l'application $s \to f^*(\alpha(g(s)))$ de G dans $\mathfrak{G}(U)$. Ceci dit, soit γ un homomorphisme algébrique de G dans $\mathfrak{G}(U)$; la classe de γ dans $\mathfrak{N}(G, U)$ correspond par l'isomorphisme précédent à un élément $g \in \mathrm{Mor}(e; G, P)$; comme g est un morphisme de la variété abélienne G dans P et comme $g(e) = 0$, g est un homomorphisme de G dans P. Les applications $s \to f^*(\alpha(g(s))) = \pi(g(s))$ et γ, qui ont même image dans $\mathfrak{N}(G, U)$, ne diffèrent que par une application constante de G dans $\mathfrak{G}(U)$; comme elles appliquent toutes deux e sur 0, elles sont égales. Il ne reste plus qu'à montrer qu'il n'y a qu'un homomorphisme g de G dans P tel que $\gamma = \pi \circ g$, i. e. que si un homomorphisme g de G dans P est tel que $\pi \circ g = 0$, on a $g = 0$. Or g est alors un élément de $\mathrm{Mor}(e; G, P)$ dont l'image dans $\mathfrak{N}(G, U)$ par l'isomorphisme considéré plus haut est nulle; on a donc bien $g = 0$, et le théorème 4 est établi.

BIBLIOGRAPHIE.

[1] P. Cartier, *Diviseurs et dérivations en géométrie algébrique*, Thèse, Université de Paris, 1958.

[2] C. Chevalley, *Fondements de la géométrie algébrique*, Cours professé à Paris; notes éditées par le Secrétariat mathématique de l'Institut Henri Poincaré, Paris, 1958.

[3] ———, *Éléments de géométrie algébrique qualitative*, à paraître prochainement.

[4] R. Godement, *Théorie des faisceaux*, Paris, 1958.

[5] J. P. Serre, "Sur la cohomologie des variétés algébriques," *Journal de Mathématiques Pures et Appliquées*, vol. 61 (1955), pp. 1-16.

[6] — ——, "Quelques propriétés des variétés abéliennes en caractéristique p," *American Journal of Mathematics*, vol. 80 (1958), pp. 715-739.

CHARACTERISTIC CLASSES AND HOMOGENEOUS SPACES, III.*

By A. Borel and F. Hirzebruch.

This paper consists of three parts, related to each other only by the fact that they bring complements to [1].

In [1, §§ 25, 26], certain expressions ($\hat{A}$-genus, Chern characters of bundles over spheres, etc.) were proved to be integers "exc 2," that is, up to a power of two. This restriction came from the fact that the proofs relied heavily on the integrality "exc 2" of the Todd genus of an almost complex manifold proved in [5]. Since then Milnor [8, 12] has shown the Todd genus to be an integer. This fact will be used in § 3 to free our earlier results from the powers of two. For this, it will be necessary to generalize slightly the notion of almost complex manifold, and to introduce between vector bundles an equivalence relation (called here S-equivalence), in which the trivial bundles form one class. These preliminaries are dealt with in §§ 1, 2.

In [1, § 23. 3], it was proved that the $\hat{A}$-genus of a coset space G/U is zero when G and U are compact, connected, semi-simple, of the same rank. The proof made use of a lemma (23. 4) stating that the sum of the positive roots of U is singular in G, which was proved essentially by case by case checking. § 4 brings an a priori proof of this lemma, in the framework of the theory of roots. When all roots of G have the same length, 23. 4 is equivalent to a theorem of de Siebenthal [10] saying that the "main diagonal" of U is singular in G. We also give a general proof of this result, which is obtained in [10] by case by case checking.

Finally, § 5 gives two elementary sufficient conditions under which the Stiefel-Whitney class $w(M)$ or the Pontrjagin class $\bar{p}(M)$ (see [1, § 9. 3]) of a compact manifold M reduces to 1, which are then applied to G/T.

The notation of [1] will be used freely.

1. S-classes of vector bundles.

1. 1. *Notation.* L stands for the field either of real numbers $\boldsymbol{R}$, or of complex numbers $\mathfrak{C}$, or of quaternions $\boldsymbol{K}$. $\boldsymbol{GL}(n, L)$ (resp. $\boldsymbol{U}(n, L)$) is the

* Received June 16, 1959.

general linear group (resp. unitary group) in L^n. A bundle with typical fibre L^n and structural group $\boldsymbol{GL}(n,L)$ or $\boldsymbol{U}(n,L)$ is called an L-vector bundle.

1.2. DEFINITION. *Let X be a topological space. Two L-vector bundles ξ, η over X are said to be S-equivalent (suspension equivalent) if there exist trivial bundles α, β such that the Whitney sums $\xi \oplus \alpha$ and $\eta \oplus \beta$ are equivalent bundles in the usual sense. The S-equivalence class, or S-class, of ξ will be denoted by $[\xi]$, and $K'(X,L)$ will be the set of S-classes of L-vector bundles over X.*

Let ξ, η be two L-vector bundles. $[\xi] = [\eta]$ means that the associated principal bundles become equivalent after the standard extension of the structural group to $\boldsymbol{U}(N,L)$, for some N, or also that the associated unit sphere bundles become equivalent after iterated suspension of the fibres.

The Whitney sum is commutative, associative, and clearly compatible with S-equivalence. Therefore it defines in $K'(X,L)$ a commutative, associative, operation for which the S-class of the trivial bundle is a zero element.

Let $f\colon X \to Y$ be a continuous map. Then, if we associate to an L-vector bundle over Y the induced bundle on X, we define clearly a homomorphism of $K'(Y,L)$ into $K'(X,L)$.

1.3. PROPOSITION. *Let X be a locally compact, paracompact, finite dimensional space. Then $K'(X,L)$ is a commutative group (with respect to Whitney sum). The S-class of the trivial bundle is the zero element.*

There remains only to show the existence of the inverse. On the Grassmann manifold $\boldsymbol{U}(n+N,L)/\boldsymbol{U}(n,L)\times \boldsymbol{U}(N,L)$ there are two canonical L-vector bundles ξ, η with typical fibres L^n, L^N, whose sum is the trivial bundle. Hence $[\xi]+[\eta]=0$. Since, by the classification theorem, any L-vector bundle with typical fibre L^n over X is induced from ξ by a map of X into the Grassmannian (for N suitably large), our assertion follows immediately.

1.4. The total Chern class of a complex vector bundle depends only on its S-class, as follows from the multiplication theorem [1, § 9.7]. It belongs to the set $\Gamma(X,\boldsymbol{Z})$ of elements of $H^*(X,\boldsymbol{Z})$ having a zero dimensional term equal to 1, and vanishing odd dimensional components. $\Gamma(X,\boldsymbol{Z})$ is a commutative group under the cup-product, and it is clear that assigning to each complex vector bundle its Chern class yields a group homomorphism of $K'(X,\boldsymbol{C})$ into $\Gamma(X,\boldsymbol{Z})$. An analogous remark can be made of course for the Pontrjagin, symplectic Pontrjagin and Stiefel-Whitney classes.

2. Weakly almost complex structures.

2.1. The standard inclusion of $\boldsymbol{GL}(n,\boldsymbol{C})$ into $\boldsymbol{GL}(2n,\boldsymbol{R})$ induces obviously a homomorphism of $K'(X,\boldsymbol{C})$ into $K'(X,\boldsymbol{R})$ to be denoted by λ. An S-class of real vector bundles is said to admit (to have) a complex structure if it belongs to the image of λ (and if an element of its inverse image has been chosen). A real vector bundle ξ admits (has) a *weak complex structure* if $[\xi]$ admits (has) a complex structure. Thus, a weak complex structure of a real vector bundle ξ is given by a trivial bundle α and a complex structure of $\xi \oplus \alpha$ in the usual sense [1, § 7.3]. Finally, a manifold is *weakly almost complex* (admits a weak almost complex structure) if its tangent bundle has been endowed with (admits) a weak complex structure.

An orientation of a real vector bundle ξ with fibre $\boldsymbol{R}^q$ is a section of the associated bundle with $\boldsymbol{O}(q)/\boldsymbol{SO}(q)$ as fibre. Since

$$\boldsymbol{O}(q)/\boldsymbol{SO}(q) \to \boldsymbol{O}(q+1)/\boldsymbol{SO}(q+1)$$

is bijective, an orientation of ξ depends only on the S-class of ξ. Thus, a weak complex structure of ξ defines an orientation of ξ. In particular, *a weakly almost complex manifold is canonically oriented.*

2.2. *Chern classes.* The Chern class of a weakly almost complex manifold X is by definition the Chern class of the weak complex structure of its tangent bundle. If X is compact, of dimension $2n$, then $c_n[X]$ is not necessarily the Euler number. For instance, take for X the unit sphere in $\boldsymbol{R}^{2n+1}$. The normal bundle and its Whitney sum with the tangent bundle are trivial. Therefore $\boldsymbol{S}_{2n}$ admits a weak almost complex structure defined by a trivial complex bundle. Then $c_n[\boldsymbol{S}_{2n}] = 0$.

2.3. *Submanifolds of codimension* 2. Let X be a compact weakly almost complex manifold, ξ its real tangent bundle, and $[\xi']$ the complex structure of $[\xi]$. Let $d \in H^2(X,\boldsymbol{Z})$. According to Thom [11] there exists an oriented submanifold D of X, of codimension 2, whose normal bundle ν is a real vector bundle with structure group $\boldsymbol{SO}(2)$ and characteristic class $i^*(d)$, where i is the embedding of D in X. Since $\boldsymbol{SO}(2) = \boldsymbol{U}(1)$, the bundle ν has a complex structure ν', whose total Chern class is $1 + i^*(d)$. Let δ be the real tangent bundle to D. Then we have in $K'(D,\boldsymbol{R})$ the equalities

$$[\delta] = [i^*\xi] - [\nu] = \lambda(i^*[\xi'] - [\nu']),$$

which show that $[\delta]$ admits a complex structure represented by a bundle δ' whose Chern class is

$$c(\delta') = i^*(c(\xi')) \cdot (1 + i^*(d))^{-1}.$$

This proves the following proposition.

2.4. PROPOSITION. *Let X be a compact weakly almost complex manifold and $c(X)$ be its total Chern class. Then every element $d \in H^2(X, \mathbf{Z})$ is representable by a submanifold D of codimension 2 which carries a weakly almost complex structure, whose total Chern class $c(D)$ is equal to $i^*(c(X) \cdot (1+d)^{-1})$, where i is the embedding of D in X.*

2.5. Let X be a compact weakly almost complex manifold of even dimension, $d \in H^2(X, \mathbf{R})$, and η a complex vector bundle over X. The Todd genus $T(X)$, the virtual Todd genus $T(d)_X$ of d, and the number $T(X, \eta)$ are then defined in exactly the same way as for an almost complex manifold. If η is a complex line bundle, with first Chern class a, then $T(X, \eta) = T(X, a)$, where $T(X, a) = T(X) - T(-a)_X$. For all this, see [5, §§ 10-12]; the definitions given there were also recalled in [1, §§ 22.1, 25.1]. It follows then from 2.4 that for every element $d \in H^2(X, \mathbf{Z})$, *the virtual Todd genus $T(d)_X$ is equal to the Todd genus of some compact weakly almost complex manifold.*

2.6. Milnor [8] (see also [12]) has established a complex analogue of cobordism theory, and has proved that the *Todd genus of a weakly almost complex manifold is an integer.* This (and 2.5) yield the

PROPOSITION. *Let X be a compact weakly almost complex manifold. Then for every $d \in H^2(X, \mathbf{Z})$, the number $T(X, d)$ is an integer.*

3. Integrality theorems for differentiable manifolds. For the definition of $\hat{A}(X, d)$ and $\hat{A}(X, d, \eta)$ we refer to [1, §§ 25.4, 25.5].

3.1. THEOREM. *Let X be a compact oriented differentiable manifold and d an element of $H^2(X, \mathbf{Z})$ whose restriction* mod 2 *is equal to $w_2(X)$. Then $\hat{A}(X, \frac{1}{2}d)$ is an integer.*

We use the notation of 25.4. Thus E/T is an almost complex manifold, $\pi: E/T \to X$ a fibre map, ξ the tangent bundle to X, and $x_1, \cdots, x_q \in H^2(E/T)$ are the roots of the Chern polynomial of a complex structure of $\pi^*(\xi)$. This implies that

$$\pi^*(w_2) = x_1 + \cdots + x_q \mod 2.$$

Furthermore, we have the equality

$$(1) \qquad \hat{A}(X, \tfrac{1}{2}d) = T(E/T, \tfrac{1}{2}(\pi^*(d) - (x_1 + \cdots + x_q))).$$

If now $d \equiv w_2 \bmod 2$, then $\pi^*(d) - (x_1 + \cdots + x_q) \equiv 0 \bmod 2$, therefore the real cohomology class $\frac{1}{2}(\pi^*(d) - (x_1 + \cdots + x_q))$ comes from an integral class under the coefficient homomorphism $\boldsymbol{Z} \to \boldsymbol{R}$. Theorem 3.1 follows then from (1) and 2.6.

3.2. COROLLARY. *Let X be a compact, oriented, differentiable manifold with vanishing second Stiefel-Whitney class. Then the genus $\hat{A}(X)$, belonging to the power series $\frac{1}{2}z^{\frac{1}{2}}/\sinh \frac{1}{2}z^{\frac{1}{2}}$, is an integer.*

In this case, we can replace d by 0 in 3.1. Since $\hat{A}(X,0) = \hat{A}(X)$, the corollary follows.

3.3. *Examples.* The polynomial $\hat{A}_1$ is equal to $-p_1/24$. Thus, if $\dim X = 4$,

$$\hat{A}(X, d/2) = ((1 + d/2 + d^2/8)(1 - p_1/24))[X] = (d^2/8 - p_1/24)[X].$$

Since $p_1[X] = 3 \cdot \tau$, where τ is the index of X (see [5, § 0.7; 11, Cor. IV.13]), we get on a 4-dimensional oriented manifold the congruence

$$d^2[X] \equiv \tau \bmod 8 \, (d \in H^2(X, \boldsymbol{Z}), d \equiv w_2 \bmod 2).$$

This can also be formulated as a statement on the quadratic form of the manifold (F. Hirzebruch-H. Hopf, *Mathematische Annalen*, vol. 136 (1958), pp. 156-172).

Let now X be 6-dimensional. Then

$$\hat{A}(X, d/2) = ((1 + d/2 + d^2/8 + d^3/48)(1 - p_1/24))[X]$$

yields the congruence

$$d^3[X] \equiv (d \cdot p_1)[X] \bmod 48, \ (d \in H^2(X, \boldsymbol{Z}), \ d \equiv w_2 \bmod 2).$$

3.4. The coefficient of p_k in $\hat{A}_k$ is $-B_k/((2k!) \cdot 2)$, where B_k is the k-th Bernouilli number [5, p. 13]. This can be easily deduced from [5, § 1]. In fact, by the usual expression of the Pontrjagin classes in terms of Chern classes [5, p. 12], $p_k = (-1)^k 2 \cdot c_{2k}$, modulo decomposable elements; since $A_k = 2^{4k} \hat{A}_k$, formula (12) of [5, p. 15] shows that the coefficient of p_k is $(-1)^k/2$ times the coefficient of c_{2k} in T_{2k}; but this coefficient is also the coefficient of $c_1{}^k$ in T_{2k} [5, Bemerkung 2, p. 15], and it follows readily from the first formula in [5, § 1.7, p. 15] that the latter is equal to $(-1)^{k-1} B_k/(2k)!$, whence our assertion. Together with 3.2, it implies:

3.5. THEOREM. *Let X be a compact oriented differentiable manifold*

of dimension $4k$ whose tangent bundle is trivial when restricted to the complement of some point in X. Then $B_k \cdot p_k[X]/((2k)!\cdot 2)$ is an integer.

This theorem has found an interesting application to the stable homotopy groups of spheres [7]. (See also [6].)

3.6. THEOREM. *Let X be a compact oriented differentiable manifold, η a complex vector bundle over X, and d an element of $H^2(X,\mathbf{Z})$ whose restriction* mod 2 *is equal to $w_2(X)$. Then $\hat{A}(X,\frac{1}{2}d,\eta)$ is an integer.*

We follow the notation of [1, § 25.5]. Thus σ is the tangent bundle to X, E/T is the total space of a bundle ξ over X with fibre map π, and $a_j \in H^2(E/T,\mathbf{Z})$ $(1 \leqq j \leqq m)$ are cohomology classes whose sum a is the first Chern class of the complex vector bundle along the fibres. Therefore a, reduced mod 2, is the second Stiefel-Whitney class of the bundle along the fibres $\hat{\xi}$. Since the tangent bundle to E/T is the sum of $\pi^*(\sigma)$ and of $\hat{\xi}$, [1, § 7.6], we have

$$\pi^*(d) + a \equiv w_2(E/T) \bmod 2. \tag{2}$$

In [1, § 25.5] it is proved that

$$\hat{A}(X,\tfrac{1}{2}d,\eta) = \sum_i \hat{A}(E/T,\tfrac{1}{2}(\pi^*(d)+2x_i+a), \qquad (x_i \in H^2(X,\mathbf{Z})).$$

Therefore, 3.6 follows from 3.1 and (2).

3.7. *Applications.* Theorem 3.6 gives a positive answer to conjecture (1) in [1, § 25.6]. Conjecture (2) ($\hat{A}(X)$ is *even* if $w_2(X) = 0$ and $\dim X \equiv 4 \bmod 8$) and also 3.6 have since been proved by a quite different method [6] which uses Bott's results [3] instead of Milnor's theorem. As a consequence, in 3.5, $B_k p_k[X]$ divided by $(2k)!\cdot 2$ is an even integer for odd k, which yields a slight sharpening of the Kervaire-Milnor theorem [7].

In [1, § 26.10] we mentioned the *theorem of Bott* [3] that the Chern class of a complex vector bundle over $\mathbf{S}_{2q}$ is divisible by $(q-1)!$, which had been proved in [1, § 25.8] only "exc 2." This and the corresponding divisibility property of Pontrjagin classes follow now from 3.6, in the same way as [1, §§ 25.8, 25.9] were derived from [1, § 25.5].

Let X be a compact almost complex manifold η a complex vector bundle over X. Then $T(X,\eta) = \hat{A}(X,\frac{1}{2}c_1,\eta)$. Therefore by 3.6, $T(X,\eta)$ *is an integer.* This gives an affirmative answer to the first question of Problem 22 in [4]. It follows also that all numbers introduced in connection with the Riemann-Roch theorem, which were proved to be integers "exc 2" in [1, § 25.6], are actually integers.

4. Some properties of roots of compact Lie groups. G will always be a compact connected semi-simple Lie group, T a maximal torus. No distinction is made between a point in the universal covering V of T and its image in T (or equivalently, between a point in the Lie algebra $\mathfrak{t}$ of T and its image in T under the exponential map); expressions like positive Weyl chamber, dominant root, simple roots are always understood with respect to some ordering. For the notation see [1, § 2].

4.1. If a, b are roots, the number $q(a,b)=2(a,b)(b,b)^{-1}$ is an integer, and $0\leqq q(a,b)\cdot q(b,a)\leqq 3$. Consequences: $q(a,b)=0,1,2,3$. If $(aa)<(bb)$ and $(ab)\neq 0$, then $q(ab)=\pm 1$. If $q(a,b)=\pm 1$, then $|q(a,b)|\leqq|q(b,a)|$, hence $(a,a)\leqq(b,b)$; if $(a,b)\neq 0$, and $(a,a)\geqq(b,b)$, then $(a,a)(b,b)^{-1}=1$, 2, 3. Let now G be simple. Then $W(G)$ is irreducible. Since the roots of a given length span a subspace invariant under $W(G)$, it follows that if G has a root of a certain length λ, then for any root a of G, there exists a root of length λ not orthogonal to a. Consequently, if the roots are normalized so that the minimal root length is one, then the other possible values are 2, 3. More precisely, it is known that the values of (a,a) are 1, 3 for $G=\boldsymbol{G}_2$, 1, 2 for $G=\boldsymbol{B}_n$, $\boldsymbol{C}_n$, $\boldsymbol{F}_4$ and 1 in the other cases. We recall that if a, b are roots, then so is $c=a-q(a,b)b$, and clearly $(a,a)=(c,c)$.

4.2. *Let G be simple, a_i $(1\leqq i\leqq l)$ be a system of simple roots, and $d=d_1a_1+\cdots+d_la_l$ be the highest root. Then d has maximal length.*

For completeness, we give a proof. There exists a positive root of maximal length $c=c_1a_1+\cdots+c_la_l$ not orthogonal to d (4.1). If $c=d$, we are done, so assume $d\neq c$. Since d is dominant, $c+d$ is not a root, therefore $(c,d)>0$, $q(c,d)=k\geqq 1$, and $c-kd$ is a root. The coefficient of a_i in $c-kd$ must then be smaller in absolute value than d_i, whence $k=1$; but then $(c,c)\leqq(d,d)$ by 4.1 and d has maximal length.

4.3. Let G be simple, a_i $(1\leqq i\leqq l)$ be simple roots of G, $d=d_1a_1+\cdots+d_la_l$ be the highest root, and U be a maximal connected semisimple subgroup containing T. Then there exists an index j such that d_j is prime, U is the centralizer of the point P_j defined by $d_j\cdot a_i(P_j)=\delta_{ij}$ $(1\leqq i\leqq l)$. The simple roots of U with respect to a suitable ordering are the a_i's $(i\neq j)$ and $-d$. The roots of U are exactly the roots of G in which a_j has coefficient 0 or $\pm d_j$. They form a *closed* system (i.e. if a, b are 2 roots of U such that $a+b$ is a root of G, then $a+b$ is a root of U). If the center of G reduces to the identity, then P_j generates a cyclic group of order d_j which is the center

of U. One obtains in this way all maximal connected semi-simple subgroups of maximal rank, up to inner automorphisms. For all this, see [2].

4.4. G being again semi-simple, let a_i $(1 \leqq i \leqq l)$ be its simple roots. Then the equations $a_1 = \cdots = a_l$ define a 1-dimensional subspace contained in the positive Weyl chamber, or also a 1-dimensional torus S in T, to be called the *main diagonal*. It belongs to a three dimensional simple group H, the *principal subgroup* of G in the sense of de Siebenthal, which is defined up to inner automorphisms by those conditions [10, § 13, Th. 2]. H is not contained in a proper subgroup of rank l [10, § 12, Th. 1].

4.5. Let $G = \boldsymbol{SU}(2)$, and Γ_n be the natural representation of degree n of G in the space of homogeneous forms of degree $n-1$ in two variables. As is known, Γ_n is up to equivalence, the only irreducible representation of degree n of G. If n is odd, it is equivalent to a real representation and not faithful. For n even, Γ_n is faithful, equivalent to the complex conjugate representation but not to a real representation. This implies in particular: if Γ is a real representation of G whose restriction to a maximal torus does not contain the trivial representation, then Γ is faithful, breaks up in a sum of real irreducible representations each of which is complex equivalent to a sum $\Gamma_n + \Gamma_n$, n even, hence $\Gamma = \Delta + \Delta$, where Δ is a sum of representations Γ_n, n even.

4.6. THEOREM. *Let U be a proper connected semi-simple subgroup of G of maximal rank. Then the main diagonal of U is singular in G* [10, § 8, Théorème 7].

We may assume that the center of G is reduced to the identity. If G is a direct product $G_1 \times G_2$, then $U = U_1 \times U_2$, where $U_i = U \cap G$ is a subgroup of maximal rank of G_i (see e.g. [2]), and its main diagonal clearly projects onto the main diagonal of U_i $(i = 1, 2)$. Using this and induction, the proof of 4.6 is easily reduced to the case where G is simple, with center reduced to (e), and U is maximal connected. Assuming this from now on, we follow the notation of 4.3 admitting moreover the simple roots to be numbered in such a way that $j = 1$. Let c^* be the point of S defined by $a_i(c^*) = 1$ $(2 \leqq i \leqq l)$ and $d(c^*) = -1$. Then

$$(1) \qquad d_1 a_1(c^*) = -1 - d_2 - \cdots - d_l.$$

$a(c^*)$ is integral for all roots a of a system of simple roots of U, hence for all roots of U; therefore, c^* is an element of the center of U.

Assume now that, contrary to our assertion, S is regular. Then $\mathrm{Ad}_{\mathfrak{g}/\mathfrak{u}}S$ does not contain the trivial representation. Let H be the principal subgroup of U containing S. By 4.5, $H=\boldsymbol{SU}(2)$, $\mathrm{Ad}_{\mathfrak{g}/\mathfrak{u}}H$ is faithful, and is a sum of two equivalent representations. From this, and from standard facts about representations of the circle group, it follows that given a complementary root a, there exists a positive complementary root $b\neq\pm a$ such that $b(s)=\pm a(s)$ for all $s\in S$. In particular, taking $a=a_1$, there exists a positive root $b=b_1a_1+\cdots+b_la_l$ not proportional to a_1, such that

$$b_1a_1(c^*)+b_2+\cdots+b_l=\pm a_1(c^*) \tag{2}$$

$(b_i\geqq 0,\ (i\geqq 1),\ b_1\geqq 1,\ (b_2,\cdots,b_l)\neq(0,\cdots,0))$. Let z be the element $\neq e$ in the center of H. Its connected centralizer U_1 in G is $\neq G$, since G has no center, and contains H, T; the last assertion of 4.4 shows then that U_1 contains U, hence is equal to U, since the latter is maximal connected. Thus, by 4.3, $d_1=2$ and $b_1=1$. It follows that in the right hand side of (2) we must have the minus sign, and we get

$$-2a_1(c^*)=b_2+\cdots+b_l, \tag{3}$$

but this, together with $b_i\leqq d_i$, obviously contradicts (1). Therefore S is singular.

4.7. There exists therefore a positive complementary root $b=b_1a_1+\cdots+b_la_l$ such that $b(s)=0$ for all $s\in S$. In particular, $b_1a_1(c^*)=-(b_2+\cdots+b_l)$ is integral; since $0<b_1<d_1$ and d_1 is prime, this and (1) show that $a_1(c^*)$ is integral. We have proved:

COROLLARY. *We keep the notation of* 4.3 *and assume d_j to be prime. Then $1+d_1+\cdots+d_l$ is divisible by d_j. If c is the linear form defined by $(a_i,c)=1$ $(i\neq j)$, $(d,c)=-1$, then (a,c) is integral for all the roots a of G.*

Before stating our next theorem, we discuss some more properties of roots.

4.8. *Let G be simple, a_i $(1\leqq i\leqq l)$ be the simple roots, and $c=c_1a_1+\cdots+c_la_l$ be a root of G. Then $c_i(a_i,a_i)\cdot(c,c)^{-1}$ is an integer $(1\leqq i\leqq l)$.*

It is known that if we perform an inversion with respect to a sphere of radius $2^{\frac{1}{2}}$ in V, then a system of roots is transformed into a system of roots (of a group G' which may or may not be isomorphic to G). Let $e\to\bar{e}$ be this transformation. Then $\bar{e}=2e\cdot(e,e)^{-1}$, and in particular

$$\bar{c} = 2 \cdot (c,c)^{-1} \sum_i c_i (a_i, a_i) \cdot 2^{-1} \cdot \bar{a}_i,$$

$$\bar{c} = \sum_i c_i ((a_i, a_i) \cdot (c,c)^{-1}) \bar{a}_i.$$

This shows first that all roots in the new system are linear combinations with coefficients of the same sign of $\bar{a}_1, \cdots, \bar{a}_l$; hence $\bar{a}_i > 0$ defines a Weyl chamber for the new system, and the $\bar{a}_i$ are a simple system of roots. Therefore the coefficients of $\bar{c}$ are integers.

4.9. *Let G be simple, U be a maximal connected subgroup of maximal rank, and b be the sum of the positive roots of U. Then (b, a) is integral for all roots a of G, the minimal root length being assumed to be* 1.

Proof. Since $W(U) \subset W(G)$, it is enough to prove this for one particular ordering. Let us consider one, say $\mathscr{S}$, with respect to which the simple roots of U are, in the notation of 4.3, $-d$ and the a_i's $(i \neq j)$. We have then [1, § 3.1]

$$(b, a_i) = (a_i, a_i) \quad (i \neq j), \quad (d, b) = -(d, d). \tag{4}$$

The minimal root length being assumed to be 1, these are all integers (4.1) and it is therefore sufficient to show that (b, a_j) is an integer. (4) yields

$$d_j(b, a_j) = -(d, d) - \sum_{i \neq j} d_i (a_i, a_i), \tag{5}$$

hence $d_j(b, a_j)$ is an integer. By 4.1, (b, a_j) is at any rate a half integer, so that we are done if d_j is odd. If all scalar products (a, a) are equal to 1, our assertion follows from (4) and 4.7; there remains therefore the case where $d_j = 2$ and (4.1) there are two root lengths. By 4.8, $d_i(a_i, a_i)(d, d)^{-1}$ is integral and by 4.2, d has maximal length; thus if $(d, d) = 2$, each term on the right hand side of (5) is even, and (b, a_j) is integral. If now $(d, d) = 3$, then $G = \mathbf{G}_2$, $d = 3a_1 + 2a_2$, which implies $j = 2$, $(a_1, a_1) = 1$ and $d_2(b, a_2) = -6$.

4.10. *Let G be simple, and assume that there are two root lengths $s < t$. Then any root of length t is the sum of two roots of length s.*

Let a be a root of length t. Since $W(G)$ is irreducible, there exists at least one root b of length s, not orthogonal to a; then $c = b - q(b, a) a$ is a root of length s (4.1). Since $q(b, a) = \pm 1$ by 4.1, our assertion is proved.

4.11. THEOREM. *Let U be a proper connected semi-simple subgroup of maximal rank of G and let b be the sum of the positive roots of U. Then b is singular in G.*

As in 4.6, it is first seen that it suffices to prove our assertion for some ordering, when G is simple and U is maximal connected.

Proof will be by contradiction. Assume that b is regular. Let then $\mathcal{S}$ be the ordering of the roots of G defined by $a > 0$ if and only if $(b, a) > 0$ [1, § 2.8]. On the roots of U, it coincides with the original ordering, with respect to which b had been defined, as follows from [1, § 3.1], hence b is also the sum of the roots of U which are positive for $\mathcal{S}$. Let us number the simple roots a_i $(1 \leqq i \leqq l)$ for $\mathcal{S}$ so that a_i is a root of U if and only if $i \leqq j$. (A priori, it is conceivable that no a_i belongs to U, in which case we set $j = 0$.) The $l - j$ other simple roots of U will then be denoted by a_i' $(j + 1 \leqq i \leqq l)$. Of course $j \neq l$ since $U \neq G$. By [1, § 3.1]

$$(6) \qquad (b, a_i) = (a_i, a_i) \quad (i \leqq j), \qquad (b, a_i') = (a_i', a_i') \quad (i \geqq j + 1).$$

For a given a_i', there exist non negative integral c_j's such that $a_i' = c_1 a_1 + \cdots + c_l a_l$, therefore

$$(7) \qquad \begin{aligned} (a_i', a_i') = (b, a_i') = c_1(a_1, a_1) + \cdots + c_j(a_j, a_j) \\ + c_{j+1}(b, a_{j+1}) + \cdots + c_l(b, a_l). \end{aligned}$$

At least two c_i's are not zero; by the definition of $\mathcal{S}$ and 4.9 we have $(a_i', a_i') \geqq 2$, hence a_i' has maximal length.

We want to prove now that $G \neq \boldsymbol{G}_2$. If G were equal to $\boldsymbol{G}_2$, then $a_i' = c_1 a_1 + c_2 a_2$ with $c_1 \cdot c_2 \neq 0$, $(a_i', a_i') = 3$, hence by 4.8 the coefficient of the root of length one would have to be a multiple of 3, but this contradicts (7) and the fact that all scalar products are integers $\geqq 1$.

Thus, $G \neq \boldsymbol{G}_2$, there are two root lengths, and $(a_i', a_i') = 2$. It also follows that two of the c_i's, say $c_{s(i)}, c_{t(i)}$ $(s(i) < t(i))$ are equal to 1, and the others to zero. Since a_i' is simple as a root of U, we must have $t(i) \geqq j + 1$, and then also $s(i) \geqq j + 1$ since the root system of U is closed (4.3); (4.8) implies then that $(a_k, a_k) = 2$ for $k = s(i), t(i)$. In particular, we see that all simple roots of G of length one belong to U.

Let now c be the first (with respect to $\mathcal{S}$) positive complementary root of length one. This exists in view of 4.10. In order to have a contradiction, it is enough to prove that $(b, c) \leqq 0$ and this will follow if we show that

$$(8) \qquad (c, a_i) \leqq 0 \quad (i \leqq j), \qquad (c, a_i') \leqq 0 \quad (i \geqq j + 1).$$

By the above, c is not a simple root, therefore $c - q(c, a_i) a_i$, expressed as linear combination of the a_i $(1 \leqq i \leqq l)$, has some positive coefficient. If $(c, a_i') \neq 0$, then $q(c, a_i') = \pm 1$ by 4.1 and because of $(c, c) = 1$, $(a_i', a_i') = 2$; since

a_i' is the sum of two simple roots, it follows again that $c - q(c, a_i')a_i'$ has some positive coefficient. Therefore, the roots $c - q(c, a_i)a_i$ $(i \leqq j)$, and $c - q(c, a_i')a_i'$ $(i \geqq j+1)$ are positive, and moreover complementary of length one since c is. By the choice of c, they must then be greater than c, in the sense of $\mathcal{S}$, and this implies (8).

5. The Stiefel-Whitney class of G/T.

5.1. Let ξ be a differentiable bundle with connected fibres. Let $b \in B_\xi$ and $F = \pi_\xi^{-1}(b)$. The normal bundle to F in E_ξ is of course trivial, since it is induced by π_ξ from the tangent space of B_ξ at b. Therefore F is orientable if E_ξ is. Furthermore, the multiplication theorem [1, § 9.7] shows that $w(F)$ (resp. $\tilde{p}(F)$) is the restriction to F of $w(E_\xi)$, (resp. $\tilde{p}(E_\xi)$). In particular, it reduces to 1 if E_ξ is parallelizable. More precisely, if the S-class (§ 1) of the tangent bundle to E_ξ is zero, then the S-class of the tangent bundle to F is also zero. A similar observation is valid for Chern classes and S-equivalence class in a complex analytic (or almost complex) bundle.

5.2. Assume now that ξ is a principal differentiable bundle. The bundle along the fibres $\hat{\xi}$ [1, § 7.4] is then parallelizable. Since the tangent bundle to E_ξ is the sum of $\hat{\xi}$ and of the bundle induced by π_ξ from the tangent bundle to B_ξ [1, § 7.6], its S-class will be zero if the S-class of the tangent bundle to B_ξ is zero. Furthermore, the multiplication theorem gives

$$\pi_\xi^*(w(B_\xi)) = w(E_\xi), \qquad \pi_\xi^*(\tilde{p}(B_\xi)) = \tilde{p}(E_\xi).$$

Hence if $w(B_\xi) = 1$ (resp. $\tilde{p}(B_\xi) = 1$), then $w(E_\xi) = 1$ (resp. $\tilde{p}(E_\xi) = 1$). If $w(E_\xi) = 1$ (resp. $\tilde{p}(E_\xi) = 1$), and π_ξ^* is injective, then $w(B_\xi) = 1$ (resp. $\tilde{p}(B_\xi) = 1$). (Coefficients in a field of characteristic two for the Stiefel-Whitney classes, arbitrary coefficients for the Pontrjagin classes.) Again a similar assertion is valid for Chern classes in the almost complex case.

5.3. PROPOSITION. *Let G be a compact connected Lie group, S a toral subgroup. Then $w(G/S) = 1$ and $\tilde{p}(G/S) = 1$.*

Let T be a maximal torus containing S. Then we have the principal fibering $(G/S, G/T, T/S)$, therefore (5.2) it is enough to prove 5.3 for $S = T$. For the Pontrjagin class, see [1, § 10.9]. There remains to prove that $w(G/T) = 1$. Without loss of generality it may be assumed that G is semi-simple and simply connected. Let Q be the subgroup of elements of order two in T. Then $(G/Q, G/T, T/Q, \pi)$ is a principal fibering. The total space, being the quotient of a group by a finite subgroup, is parallelizable, therefore,

(5.2) it will be enough to show that π^* is injective in cohomology mod 2. Since G and G/T are simply connected, $\pi_1(G/Q) = Q$, and the map $\pi_1(T/Q) \to \pi_1(G/Q)$ defined by the inclusion i is surjective. It follows easily that $i^*\colon H^*(G/Q, \mathbf{Z}_2) \to H^*(T/Q, \mathbf{Z}_2)$ is an isomorphism in dimension 1. But T/Q is a torus, hence $H^*(T/Q, \mathbf{Z}_2)$ is generated by its element of degree $\leqq 1$; therefore i^* is surjective, the fibre is totally non homologous to zero in cohomology mod 2; as is well known, this implies that π^* is injective.

5.4. It can also be derived directly from 5.1, 5.2 that *the S-class* (§ 1) *of the tangent bundle to G/S (S toral subgroup of G) is zero.*

In view of 5.2 and of the existence of the principal fibering $(G/S, G/T, T/S)$ it is enough to prove 5.4 when $S = T$ is a maximal torus. Let $\mathfrak{g}$ be the Lie algebra of G, $\mathfrak{R}$ the set of regular elements, and let G operate on $\mathfrak{g}$ by the adjoint representation. Since the centralizer of a regular element $x \in \mathfrak{g}$ is the maximal torus containing the one-parameter subgroup spanned by x, the orbits of G in $\mathfrak{R}$ are homeomorphic to G/T. Moreover, it is classical, and easily checked, that these orbits are the fibres in a differentiable fibering of $\mathfrak{R}$. Since $\mathfrak{R}$ is parallelizable, as an open subset of $\mathfrak{g}$, our assertion follows from 5.1.

The nullity of $w_2(G/T)$ was noticed in [1, § 22.3] and, as remarked above, $\bar{p}(G/T) = 1$ was also proved in [1, § 10.9].

5.5. Without entering into details, let us mention a case containing the preceding one, in which 5.1 applies. Let G operate differentiably on a connected manifold M. Among the different stability groups $G_x = \{g \in G, g \cdot x = x\}$, let H be one of smallest dimension, which has the minimal number of connected components among stability groups of that dimension. Then the set of points whose stability group is conjugate to H is an open set in M, which is differentiably fibered by the orbits [9, pp. 221-222]. Those are homeomorphic to G/H, and are called the *main orbits.* 5.1 yields then the

PROPOSITION. *Let G be a compact Lie group acting on a connected manifold M, and let F be a main orbit. Then F is orientable if M is. $w(F)$ and $\bar{p}(F)$ are the restrictions to F of $w(M)$ and $\bar{p}(M)$. If the S-class of the tangent bundle to M is zero, then the S-class of the tangent bundle to F is zero. In particular, if G/H is homeomorphic to the main orbit of a linear representation then it is orientable, the S-class of its tangent bundle is zero, and $w(G/H) = 1$, $\bar{p}(G/H) = 1$.*

The proof given in 5.4 is the particular case of 5.5 corresponding to the adjoint representation, where the main orbits are homeomorphic to G/T.

INSTITUTE FOR ADVANCED STUDY,
UNIVERSITY OF BONN, GERMANY.

REFERENCES.

[1] A. Borel and F. Hirzebruch, "Characteristic classes and homogeneous spaces," Part I, *American Journal of Mathematics*, vol. 80 (1958), pp. 459-538, Part II, *ibid.*, vol. 81 (1959), pp. 315-382.

[2] ——— and J. de Siebenthal, "Sur les sous-groupes fermés connexes des groupes de Lie clos," *Commentarii Mathematici Helvetici*, vol. 23 (1949), pp. 200-221.

[3] R. Bott, "The space of loops on a Lie group," *The Michigan Mathematical Journal*, vol. 5 (1958), pp. 35-61.

[4] F. Hirzebruch, "Some problems on differentiable and complex manifolds," *Annals of Mathematics*, vol. 60 (1954), pp. 213-236.

[5] ———, "Neue topologische Methoden in der algebraischen Geometrie," *Ergebnisse der Mathematik und ihrer Grenzgebiete* (*N. F.*) 9, Berlin, 1956.

[6] ———, "A Riemann-Roch theorem for differentiable manifolds," *Séminaire Bourbaki*, Exp. 177, Février 1959.

[7] M. Kervaire and J. Milnor, "Bernoulli numbers, homotopy groups and a theorem theorem of Rohlin," *Proceedings of the International Congres of Mathematicians*, Edinburgh, 1958 (to appear).

[8] J. Milnor, "On the cobordism ring, and a complex analogue," *American Journal of Mathematics*.

[9] D. Montgomery-L. Zippin, *Topological transformation groups*, New York, 1955.

[10] J. de Siebenthal, "Sur les sous-groupes fermés connexes des groupes de Lie clos," *Commentarii Mathematici Helvetici*, vol. 25 (1951), pp. 210-256.

[11] R. Thom, "Quelques propriétés globales des variétés différentiables," *ibid.*, vol. 28 (1954), pp. 17-85.

[12] ———, "Travaux de Milnor sur le cobordisme," *Séminaire Bourbaki*, Exp. 180, Février 1959.

ON THE COBORDISM RING Ω^* AND A COMPLEX ANALOGUE, PART I.*

By J. MILNOR.

This paper will prove that the cobordism groups Ω^i, defined by Thom [15], have no odd torsion.[1] Furthermore, it is shown that certain related groups $\pi_{i+2n}M(U_n)$ have no torsion at all; providing that n is large. The proofs are based on a spectral sequence due to J. F. Adams [1, 2].

The following is a brief summary of Thom's constructions. Let G be a subgroup of the orthogonal group O_n. (More generally one could start with any Lie group G, together with a specified representation into O_n.) Beginning with a universal bundle for G we can form:

1) The weakly associated bundle having the disk D^n as fibre. Let $\pi\colon E \to B(G)$ denote the projection map of this bundle.

2) The weakly associated bundle having the sphere S^{n-1} as fibre. Let $\partial E \subset E$ denote the total space.

The *Thom space* $M(G)$ is now defined as the identification space obtained from E by collapsing ∂E to a point.

Taking G to be the rotation group $SO_n \subset O_n$, Thom showed that the homotopy group $\pi_{i+n}M(SO_n)$ is independent of n, providing that n is large. He showed that this group is isomorphic to the "cobordism group" Ω^i; and determined its structure up to torsion. The 2-torsion subgroup of Ω^i has recently been determined by C. T. C. Wall. Hence the assertion that Ω^i has no odd torsion completes the description of this group.

Let $M(U_n)$ denote the Thom space for the unitary group $U_n \subset O_{2n}$. In Part II of this paper it will be shown that the stable homotopy group $\pi_{i+2n}M(U_n)$ can be interpreted as a "complex cobordism group." Part I will determine the structure of this group without attempting to interpret it.

* Received July 27, 1959.

[1] Added in proof. This result has been obtained independently by B. G. Averbuch, *Doklady Akademii Nauk SSSR*, vol. 125 (1959), pp. 11-14. The results on complex cobordism have been obtained independently by Novikov.

The first section proves several lemmas concerning the Steenrod algebra, which are needed later. The second section describes the Adams spectral sequence, which relates the cohomology module of any space to the stable homotopy groups of the space. Sections 3 and 4 complete the argument by computing the cohomology modules of $M(U_n)$ and $M(SO_n)$ respectively.

1. Lemmas concerning the Steenrod algebra. Let A denote the Steenrod algebra corresponding to a fixed prime p. (See Cartan [6], Adem [3].) The Bochstein coboundary operation will be denoted by $Q_0 \in A^1$. The two-sided ideal generated by Q_0 in A will be denoted by (Q_0).

LEMMA 1. *The Steenrod algebra contains a subalgebra A_0 with the following properties.*

(i) *A_0 is a Grassmann algebra over Z_p with generators $Q_0, Q_1, \cdots$ of odd dimension.*

(ii) *A is free as a right A_0-module.*

(iii) *The identity map of A induces an isomorphism between the left A-modules $A \otimes_{A_0} Z_p$ and $A/(Q_0)$.*

[*Explanation of* (iii). The field Z_p is considered as a left A_0-module with $Q_i Z_p = 0$. Hence $A \otimes_{A_0} Z_p$ is the quotient of A by the left ideal $AQ_0 + AQ_1 + AQ_2 + \cdots$.]

Proof for the case p odd. We will first prove the corresponding statements with left and right interchanged. According to Milnor [10, Theorem 4a]:

(1) There is a basis for A over Z_p consisting of elements $Q_0^{e_0} Q_1^{e_1} \cdots \mathcal{P}^R$. Here the integers $e_0, e_1, \cdots$ should be 0 or 1, and almost all zero. The letter R stands for a sequence $(r_1, r_2, \cdots)$ of non-negative integers, almost all zero.

[*Explanation.* The element $\mathcal{P}^R$ is a complicated polynomial in the Steenrod operations, with dimension $\sum r_j(2p^j - 2)$. For the special case $R = (r, 0, 0, 0, \cdots)$ the element $\mathcal{P}^R$ is equal to the Steenrod operation $\mathcal{P}^r$. The element Q_i of dimension $2p^i - 1$ can be defined inductively by the rule $Q_{i+1} = \mathcal{P}^{p^i} Q_i - Q_i \mathcal{P}^{p^i}$.]

Furthermore:

(2) The elements Q_i are odd dimensional, and satisfy $Q_i Q_j + Q_j Q_i = 0$, $Q_i Q_i = 0$.

Thus the Q_i generate a Grassmann algebra which may be denoted by $A_0 \subset A$. Clearly A is free as a left A_0-module, with basis $\{\mathcal{P}^R\}$.

Consider the right ideal $Q_0A + Q_1A + Q_2A + \cdots$. The following identity (see [10, Theorem 4a]) proves that this is also a left ideal. Define $p^i\Delta_j$ as the sequence $(0, \cdots, 0, p^i, 0, \cdots)$ with p^i in the j-th place.

(3) $\mathcal{P}^R Q_i$ is equal to $Q_i \mathcal{P}^R + \sum Q_{i+j} \mathcal{P}^{R-p^i\Delta_j}$, to be summed over all $j > 0$ for which $R - p^i\Delta_j$ is a sequence of non-negative integers. (That is, all j for which $r_j \geqq p^i$.)

Thus $Q_0A + Q_1A + \cdots$ is a two-sided ideal which contains Q_0, and therefore contains (Q_0).

As a special case of (3), the identity $\mathcal{P}^{\Delta_j} Q_0 = Q_0 \mathcal{P}^{\Delta_j} + Q_j$ is valid. Thus the elements Q_j belong to the ideal (Q_0). This proves that the ideal $Q_0A + Q_1A + \cdots$ is equal to (Q_0). Dividing A by these ideals, it follows that $Z_p \otimes_{A_0} A$ is isomorphic to $A/(Q_0)$.

This proves Lemma 1 for p odd, except that right and left have been interchanged. To complete the proof it is only necessary to recall:

(4) There exists an anti-automorphism of A; that is, a Z_p-isomorphism $c \colon A \to A$ satisfying

$$c(xy) = (-1)^{\dim x \dim y} c(y) c(x).$$

Furthermore, c carries Q_i into $-Q_i$.

This is proved in [10, § 7]. Clearly Lemma 1 follows (for p odd).

Lemma 2. *The elements $\mathcal{P}^R \in A$ yield a basis over Z_p for the quotient algebra $A/(Q_0)$.*

Proof for p odd. Recall that $\{\mathcal{P}^R\}$ forms a basis for A, considered as a left A_0-module. Hence it forms a basis for $Z_p \otimes_{A_0} A = A/(Q_0)$ over Z_p, which completes the proof.

Conventions. The sum $R + R'$ of two sequences is defined as the term by term sum, and nR denotes the sequence $(nr_1, nr_2, \cdots)$. The binomial coefficient (R, R') is defined as the product over i of $(r_i + r'_i)!/r_i!r'_i!$. The symbol Δ_j stands for a sequence with 1 in the j-th place and zero elsewhere.

Proof of Lemmas 1 *and* 2 *for the case $p = 2$.* The Steenrod algebra over Z_2 has a basis consisting of elements Sq^R of dimension $r_1 + 3r_2 + 7r_3 + \cdots$. (See [10, Appendix 1].) Define $\mathcal{P}^R$ to be Sq^{2R} and define Q_{i-1} to be Sq^{Δ_i}. (For example $Q_0 = \mathrm{Sq}^{\Delta_1} = \mathrm{Sq}^1$ which checks with the definition of Q_0 as the

Bochstein coboundary operator.) Then we will prove Assertions (1), (2), (3) and (4) above. Using these, the proof of Lemmas 1 and 2 can be carried out just as for p odd.

The formula for products $\mathrm{Sq}^{R'}\mathrm{Sq}^{R}$ is rather complicated; however the following special case will suffice.

(5) If E is a sequence satisfying $e_i \leqq 1$, then $\mathrm{Sq}^{E}\mathrm{Sq}^{R}$ is equal to $(E, R)\,\mathrm{Sq}^{E+R}$.

For a proof see [10, Corollary 4 and Appendix 1]. As examples, taking $E = \Delta_{i+1}$, $R = \Delta_{j+1}$ we find that $Q_iQ_j = Q_jQ_i$, and that $Q_iQ_i = 0$. This proves Assertion (2) for the case $p = 2$.

By induction the product $Q_0^{e_1}Q_1^{e_2}\cdots$ is equal to Sq^{E}. Furthermore, a binomial coefficient of the form $(E, 2R)$ is always odd, hence $\mathrm{Sq}^{E}\mathcal{P}^{R} = \mathrm{Sq}^{E}\mathrm{Sq}^{2R}$ is equal to Sq^{E+2R}. Since every sequence can be written uniquely in the form $E + 2R$, it follows that these elements form a basis for A over Z_2. This proves Assertion (1).

Proof of Assertion (3) *for* $p = 2$. Direct application of the general product rule [10, Theorem 4b] shows that

$$\mathrm{Sq}^{2R}\mathrm{Sq}^{\Delta_{i+1}} = \mathrm{Sq}^{\Delta_{i+1}}\mathrm{Sq}^{2R} + \sum \mathrm{Sq}^{2R-2^{i+1}\Delta_j+\Delta_{i+1+j}},$$

to be summed over all $j \geqq 1$ for which $r_j \geqq 2^i$. On the other hand, using Assertion (5), the j-th term on the right can be written as

$$\mathrm{Sq}^{\Delta_{i+1+j}}\mathrm{Sq}^{2R-2^{i+1}\Delta_j} = Q_{i+j}\mathcal{P}^{R-2^i\Delta_j}.$$

Thus $\mathcal{P}^{R}Q_i = Q_i\mathcal{P}^{R} + \sum Q_{i+j}\mathcal{P}^{R-2^i\Delta_j}$, as required.

Since Assertion (4) is also true for $p = 2$, this completes the proof of Lemmas 1 and 2.

[*Remark.* There is one essential difference between the case p odd and the case $p = 2$. For p odd the elements $\mathcal{P}^{R}$ span a subalgebra of A isomorphic to $A/(Q_0)$; but for $p = 2$ there is no such subalgebra. This can be seen using the identity $\mathrm{Sq}^2\mathrm{Sq}^2 = \mathrm{Sq}^1\mathrm{Sq}^2\mathrm{Sq}^1 \neq 0$.]

The symbol Δ_0 will denote the sequence $(0, 0, \cdots)$.

LEMMA 3. *If p is odd, then the cohomology operations $\mathcal{P}^{R}$ have the following properties.*

(1) *For $x, y \in H^*(X; Z_p)$ the element $\mathcal{P}^{R}(xy)$ is equal to*

$$\sum_{R_1+R_2=R} (\mathcal{P}^{R_1}x)(\mathcal{P}^{R_2}y).$$

(2) *For a 2-dimensional cohomology class $t \in H^2(X; Z_p)$, the element $\mathcal{P}^R t$ is equal to t^{p^i} if $R = \Delta_i$; and is zero if R is not equal to one of the sequences $\Delta_0, \Delta_1, \Delta_2, \cdots$.*

Proof. The first assertion follows from [10, Lemma 9]. For the special case $R = r\Delta_1$, the second assertion is well known. That is:

$$\mathcal{P}^0 t = t, \quad \mathcal{P}^1 t = t^p, \quad \mathcal{P}^r t = 0 \text{ for } r > 1.$$

But every $\mathcal{P}^R$ is a "polynomial" in the Steenrod operations $\mathcal{P}^r$. Proceeding by induction on the complexity of this polynomial, we see that $\mathcal{P}^R t$ must have the form kt^i, where $k \in Z_p$ is some constant, and $2i$ is the dimension.

To evaluate k it is sufficient to consider one example. As example, let X be the $2i$-skeleton of the Eilenberg-MacLane complex $K(Z_p, 1)$. According to [10, Lemmas 4, 6] we have:

$$\lambda(t) = t \otimes \xi_0 + t^p \otimes \xi_1 + \cdots;$$

hence

$$\mathcal{P}^R t = \sum_i \langle \mathcal{P}^R, \xi_i \rangle t^{p^i}.$$

Using the definition of $\mathcal{P}^R$, this is equal to t^{p^i} if $R = \Delta_i$ and is zero otherwise. This completes the proof.

For the prime $p = 2$, both assertions of Lemma 3 would be false. However the following modified assertions are proved by the same method:

(1′) $\mathrm{Sq}^R(xy) = \sum_{R_1 + R_2 = R} (\mathrm{Sq}^{R_1} x)(\mathrm{Sq}^{R_2} y)$.

(2′) If $a \in H^1(X; Z_2)$, then $\mathrm{Sq}^{\Delta_i} a = a^{2^i}$; and $\mathrm{Sq}^R a = 0$ for R not of the form Δ_i.

Using these statement the following result will be proved.

LEMMA 3′. *Let $p = 2$ and let $H^*(X; Z_2)$ be a cohomology ring which is annihilated by the operation $Q_0 = \mathrm{Sq}^1$. The assertions* (1) *and* (2) *of Lemma* 3 *are valid as originally stated.*

Proof of (1). If R_1 is a sequence containing some odd integer, then Sq^{R_1} belongs to the ideal (Q_0) (compare the proof of Lemma 1), and therefore annihilates the cohomology of X. Thus in formula (1′) above, it is sufficient to consider sequences R_1 and R_2 which are "even." This proves assertion (1).

Proof of (2). It will be convenient to weaken the hypothesis on X, and assume only that $\mathrm{Sq}^1 t = 0$. Then just as in the proof of Lemma 3, it follows

that $\mathfrak{P}^R t$ has the form kt^i. In order to determine the constant $k \in Z_2$, it is sufficient to consider the example of a real projective space X, with $t = a^2$, Using (1′) and (2′) it is seen that $\mathfrak{P}^R t$ equals t^{2^i} for $R = \Delta_i$ and equals zero otherwise. This completes the proof of Lemma 3′.

2. **The spectral sequence of Adams.** Let X, Y be finite CW-complexes with base point denoted by o; and let A be the Steenrod algebra for some fixed prime p. Thus the cohomology group $H^*(X \bmod o; Z_p)$ is a graded left A-module.

The m-fold *suspension* $S^m X$ is obtained from the product $X \times I^m$ by collapsing $(X \times \partial I^m) \cup (o \times I^m)$ to a point. Here I^m denotes the unit m-cube. The *stable track group* $\{X, Y\}_n$ is the direct limit under suspension of the group of homotopy classes of maps $S^{m+n} X \to S^m Y$. (The integer n may be positive or negative.)

THEOREM OF ADAMS. *There exists a spectral sequence $\{E_r^{st}, d_r\}$ determined by X, Y and p such that*

$$E_2^{st} = \mathrm{Ext}_A^{st}(H^*(Y \bmod o; Z_p), H^*(X \bmod o; Z_p))$$

and such that

$$E_\infty^{st} = B^{st}/B^{s+1\, t+1},$$

where $\{X, Y\}_n = B^{0\, n} \supset B^{1\, n+1} \supset B^{2\, n+2} \supset \cdots$ is a certain filtration. The intersection $\bigcap_s B^{s\, n+s}$ of these groups is equal to the subgroup of $\{X, Y\}_n$ consisting of elements whose order is finite and prime to p. Each succeeding term E_{r+1} of the spectral sequence is equal to the homology of E_r with respect to the differential operator

$$d_r \colon E_r^{st} \to E_r^{s+r\, t+r-1};$$

and E_∞ is the limit as $r \to \infty$ of E_r.

The functor Ext_A^{st} is defined as follows. If M and N are graded left A-modules let $\mathrm{Hom}_A{}^t(M, N) = \mathrm{Ext}_A^{0t}(M, N)$ denote the group of A-homomorphisms $M \to N$ of degree $-t$. Choose a projective resolution

$$\cdots \to P_2 \xrightarrow{d} P_1 \xrightarrow{d} P_0 \to M \to 0,$$

where the A-homomorphisms d have degree zero. Then $\mathrm{Ext}_A^{st}(M, N)$ is defined as the homology group (kernel modulo image) of the sequence

$$\mathrm{Hom}_A{}^t(P_{s-1}, N) \xrightarrow{d^*} \mathrm{Hom}_A{}^t(P_s, N) \xrightarrow{d^*} \mathrm{Hom}_A{}^t(P_{s+1}, N).$$

It will be convenient to add an E_1 term to the spectral sequence by defining $E_1^{st} = \mathrm{Hom}_A{}^t(P_s, N)$, $d_1 = d^*$.

For the special case $X = S^0$ this theorem is proved in Adams [1]. The more general case is proved by the same argument. It is only necessary to replace the homotopy group $\{S^0, \}_n$ by the track group $\{X, \}_n$ throughout. See Adams [2].

More generally the finite complex Y may be replaced by a "spectrum" in the sense of Lima [9] and Spanier [13]; or by an "object in the stable category" in the sense of Adams [2]. For our purpose the following definition will be convenient. A *stable object* $\boldsymbol{Y}$ is a sequence of CW-complexes $(Y_0, Y_1, \cdots)$ such that each suspension SY_i is a subcomplex of Y_{i+1}. The imbedding $SY_i \subset Y_{i+1}$ must be explicitly given.

Given such an object, define the chain group $C_n(\boldsymbol{Y})$ as the direct limit under suspension of the chain groups $C_{n+i}(Y_i \bmod o)$. Homology and cohomology groups are then defined as usual. Similarly, for any finite complex X define $\{X, \boldsymbol{Y}\}_n = \text{dir.lim.}\{S^iX, Y_i\}_n$. The abbreviation $\pi_n\boldsymbol{Y}$ will sometimes be used for $\{S^0, Y\}_n$.

Remark. The suspension homomorphism of chain groups should be defined by the correspondence

$$\alpha \to \alpha \times \iota, \text{ for } \alpha \in C_*(Y_i \bmod o),\ \iota \in C_1(I \bmod \partial I),$$

so as to commute with boundary homomorphisms.

Examples. Any finite complex Y may be defined with the stable object

$$\boldsymbol{Y} = (Y, SY, S^2Y, \cdots).$$

We will see later that the suspension of the Thom space $M(SO_n)$ is imbedded naturally as a subcomplex of $M(SO_{n+1})$. Hence the *stable Thom object*

$$\boldsymbol{M}(SO) = (o, M(SO_1), M(SO_2), \cdots)$$

is defined. Note that the track group

$$\{S^0, \boldsymbol{M}(SO)\}_n = \text{dir. lim. } \pi_{n+i}(M(SO_i))$$

is isomorphic to the cobordism group Ω^n.

Assertion. The theorem of Adams remains valid if the finite complex Y is replaced by any stable object $\boldsymbol{Y}$; providing that the following *finiteness condition* is satisfied. The groups $C_n(\boldsymbol{Y}; Z)$ should be finitely generated, and should vanish for n less than some constant.

This can be proved in two ways. One can simply take the direct limit of the spectral sequences for the "finite sub-objects" of $\mathbf{Y}$; or the theorem can be proved from the beginning in the stable category. See Adams [2]. The second approach is preferable, since the proof is much easier in the stable category. Details will not be given.

Using the Adams spectral sequence we will prove the following key result. Let $\mathbf{Y}$ be an object such that $H^n(\mathbf{Y};Z_p)$ is zero for n odd. Then $H^*(\mathbf{Y};Z_p)$ is annihilated by the element Q_0, and hence can be considered as a graded module over the quotient algebra $A/(Q_0)$.

THEOREM 1. *If $H^*(\mathbf{Y};Z_p)$ is a free $A/(Q_0)$-module with even dimensional generators, and if $C_*(\mathbf{Y};Z)$ satisfies the finiteness condition, then the stable homotopy group $\{S^0,\mathbf{Y}\}_*$ contains no p-torsion.*

The idea of the proof is to compute the spectral sequence for the track group $\{X,\mathbf{Y}\}_*$, where X is a "co-Moore space" having cohomology groups $H^i(X \bmod o;Z)$ equal to Z_p for $i=k$ and equal to zero for $i\neq k$.

The following universal coefficient theorem has been proved by Peterson [11]. *There exists an exact sequence*

$$0\to\{S^k,\mathbf{Y}\}_n\otimes Z_p\to\{X,\mathbf{Y}\}_n\to\operatorname{Tor}(\{S^k,\mathbf{Y}\}_{n-1},Z_p)\to 0.$$

An immediate consequence is the following.

LEMMA 4. *If $\{S^0,\mathbf{Y}\}_*$ contains p-torsion, then $\{X,\mathbf{Y}\}_m$ must be non-trivial for two consecutive values of m.*

On the other hand, assuming that $H^*(\mathbf{Y};Z_p)$ is a free $A/(Q_0)$-module on even dimensional generators, we will see that $\{X,\mathbf{Y}\}_m$ is zero for m odd. This will prove Theorem 1.

Construction of an A-free resolution for $H^*(\mathbf{Y};Z_p)$.

First consider the Grassmann algebra A_0 and the A_0-module Z_p. According to Cartan's theory of constructions, to each Grassmann algebra A_0 there corresponds a twisted polynomial algebra P and a differential operator d on $A_0\otimes P$ so that this tensor product becomes acyclic. If A_0 has generators $Q_0, Q_1,\cdots$, then P has a basis over Z_p consisting of elements $b(r_0,r_1,\cdots)$ of dimension $\sum r_i(\dim Q_i+1)$. The integers $r_0,r_1,\cdots$ should be non-negative and almost all zero. The differential operator d is defined as follows. (In order to make the signs come out correctly, we let d act on the right.) For any $a\in A_0$:

$$a \otimes b(r_0, r_1, \cdots) d = \sum a Q_i \otimes b(r_0, \cdots, r_i - 1, r_{i+1}, \cdots),$$

summed over all i for which $r_i > 0$.

Proof that $A_0 \otimes P$ is acyclic. For a Grassmann algebra on one generator, see Cartan [5, p. 704, I]. But a Grassmann algebra with finitely many generators in each dimension can be considered as a tensor product of Grassmann algebras with one generator. Hence the conclusion follows by applying the Künneth theorem.

This conclusion can be formulated as follows. Let F_s be the free A_0-module generated by those symbols $b(r_0, r_1, \cdots)$ for which $r_0 + r_1 + \cdots = s$. Then $A_0 \otimes P$ can be considered as the direct sum $F_0 + F_1 + \cdots$. The augmentation $\epsilon\colon F_0 \to Z_p$ is the A_0-homomorphism defined by $b(0, 0, \cdots) \to 1$. It follows that *the sequence*

$$\cdots \xrightarrow{d} F_1 \xrightarrow{d} F_0 \xrightarrow{\epsilon} Z_p \to 0$$

is an A_0-free resolution of Z_p.

Now apply the functor $A \otimes_{A_0}$ to this exact sequence. Since A is free as a right A_0-module, we obtain an exact sequence

$$\cdots \to A \otimes_{A_0} F_1 \to A \otimes_{A_0} F_0 \to A \otimes_{A_0} Z_p \to 0$$

of left A-modules. Furthermore, each $A \otimes_{A_0} F_s$ is a free A-module. Thus we have constructed an A-free resolution of $A \otimes_{A_0} Z_p$.

According to Lemma 1, the A-module $A/(Q_0)$ is isomorphic to $A \otimes_{A_0} Z_p$. Hence in order to form an A-free resolution of any $A/(Q_0)$-free module, it is sufficient to take the direct sum of a number of copies of the above resolution. This proves the following.

LEMMA 5. *Let $H^*(\mathbf{Y}; Z_p)$ be a free module over $A/(Q_0)$ with basis $\{y_\alpha\}$. Then there exists an A-free resolution*

$$\cdots \to F_1' \to F_0' \to H^*(\mathbf{Y}; Z_p) \to 0,$$

where each F_s' has a basis consisting of elements $b_\alpha(r_0, r_1, \cdots)$, with $r_0 + r_1 + \cdots = s$. The dimension of such a basis element is equal to $\dim y_\alpha + \sum 2r_i(p^i - 1) + s$.

[*Explanation.* The integer s has been added to the dimension of $b_\alpha(r_0, r_1, \cdots)$ so that the homomorphisms $d'\colon F_s' \to F_{s-1}'$ will have degree zero.]

Now consider the complex X consisting of a circle with a 2-cell attached by a map of degree p. Let

$$x \in H^1(X \bmod o; Z_p), \qquad Q_0 x \in H^2(X \bmod o; Z_p)$$

be generators. Then the term

$$E_1^{st} = \operatorname{Hom}_A{}^t(F_s', H^*(X \bmod o; Z_p))$$

of the spectral sequence for $\{X, \boldsymbol{Y}\}$ has a basis consisting of the following elements.

(1) For each $b_\alpha(r_0, r_1, \cdots)$ of dimension $t+1$, the homomorphism $h_\alpha(r_0, r_1, \cdots)$ which carries this basis element into x and carries the other basis elements into zero.

(2) For each $b_\alpha(r_0, r_1, \cdots)$ of dimension $t+2$, the homomorphism $h_\alpha'(r_0, r_1, \cdots)$ which carries this basis element into $Q_0 x$ and carries the other basis elements into zero.

The boundary operator $d_1\colon E_1^{st} \to E_1^{s+1\, t}$ is given by

$$d_1 h_\alpha(r_0, r_1, \cdots) = h_\alpha'(r_0 + 1, r_1, \cdots),$$

and

$$d_1 h_\alpha'(r_0, r_1, \cdots) = 0.$$

Thus E_2^{st} has as basis the set of elements $h_\alpha'(0, r_1, r_2, \cdots)$, with total dimensions $t - s$ equal to $\dim y_\alpha + \sum 2r_i(p^i - 1) - 2$.

If the integers $\dim y_\alpha$ are all even, then everything in the spectral sequence is even dimensional. It follows that $\{X, \boldsymbol{Y}\}_m$ is zero for m odd. Together with Lemma 4, this completes the proof of Theorem 1.

3. Computation of $H^*(B(U_n); Z_p)$ and $H^*(\boldsymbol{M}(U); Z_p)$. This section will complete the study of $M(U_n)$ by constructing a stable object

$$\boldsymbol{M}(U) = (o, o, M(U_1), SM(U_1), M(U_2), SM(U_2), \cdots);$$

and showing that $H^*(\boldsymbol{M}(U); Z_p)$ is a free module over $A/(Q_0)$, with even dimensional generators, for any prime p.

The proof of this assertion is an immediate generalization of the argument which Thom used to compute the non-orientable cobordism group. In our terminology, Thom showed that $H^*(\boldsymbol{M}(O); Z_2)$ is a free A-module. (See [15, pp. 39-42].)

First a description of $H^*B(U_n)$. *The coefficient group Z_p is to be*

understood, where p is some fixed prime. (However, integer coefficients could equally well be used.) Let $T_n \subset U_n$ be the n-torus consisting of diagonal unitary matrices. There is a natural map $B(T_n) \to B(U_n)$ of classifying spaces. The cohomology algebra $H^*B(T_n)$ is a polynomial algebra on generators $t_1, \cdots, t_n$ of dimension 2. According to Borel and Serre [4] we may identify $H^*B(U_n)$ with the subalgebra consisting of all symmetric polynomials.

A basis for $H^{2r}B(U_n)$ over Z_p is given as follows. Let $\omega = i_1 \cdots i_k$ range over all partitions of r such that the "length" k is less than or equal to n. (A partition of r is an unordered sequence of positive integers with sum r.) *Define* $s(\omega)$ as the "smallest" symmetric polynomial which contains the term $t_1^{i_1} \cdots t_k^{i_k}$.

[The notation $\sum t_1^{i_1} \cdots t_k^{i_k}$ is commonly used. A more precise definition would be the following. Consider all distinct monomials which can be obtained from $t_1^{i_1} \cdots t_k^{i_k}$ by permuting the n variables; and let $s(\omega)$ denote their sum. It is clear that these elements $s(\omega)$ from a basis for the vector space of symmetric polynomials.]

Next we must study the Thom complex $M(U_n)$. For a group $G \subset SO_m$ recall that $M(G)$ is the quotient space $E/\partial E$, where E is an oriented m-disk bundle over $B(G)$. Any CW-cell subdivision of $B(G)$ induces a cell subdivision of $M(G)$ as follows. For each open i-cell e of $B(G)$, the inverse image e' in $E - \partial E$ is an $(i+m)$-cell. Clearly, $M(G)$ is the disjoint union of these cells e', together with the base point. It is not difficult to verify that $M(G)$ thus becomes a CW-complex.

Let $G \times 1$ denote the group G, considered as a subgroup of SO_{m+1}. The CW-complex $M(G \times 1)$ can be identified with the suspension $SM(G)$ as follows. Let D^m denote the m-disk and I the unit interval. Map $D^m \times I$ onto D^{m+1} by the correspondence

$$(x_1, \cdots, x_m), y \to x_1, \cdots, x_m, (2y-1)(1-x_1^2-\cdots-x_m^2)^{\frac{1}{2}}.$$

This correspondence gives rise to a map f of $E \times I$ onto the total space E_1 of the associated $(m+1)$-disk bundle. Since f carries $(\partial E \times I) \cup (E \times \partial I)$ onto the boundary ∂E_1, it follows that f gives rise to a map f': $SM(G) \to M(G \times 1)$. But f is a relative homeomorphism, hence f' is a homeomorphism.

The Thom isomorphism

$$\phi\colon H^iB(G) \to H^{i+m}(M(G) \bmod o)$$

is defined as follows. (see [14, Théorème I.4]). The cohomology of $M(G) \bmod o$ will be identified with the cohomology of $E \bmod \partial E$. It can be

verified that $H^m(E \bmod \partial E; Z)$ is an infinite cyclic group, with standard generator u. The isomorphism ϕ is now defined by the formula $\phi(a) = \pi^*(a)u$, where $\pi\colon E \to B(G)$ denotes the projection map. It follows from this definition that the following diagram is commutative:

$$\begin{array}{ccc} H^{i+m}(M(G) \bmod o) & \xrightarrow{S} & H^{i+m+1}(M(G \times 1) \bmod o) \\ \uparrow \phi & & \uparrow \phi \\ H^iB(G) & = & H^iB(G \times 1). \end{array}$$

Here S denotes the cohomology suspension, defined using the cohomology cross product.

Now let us specialize to the case $G = U_n \subset SO_{2n}$. The classifying space $B(U_n)$ has a standard cell subdivision due to Ehresmann [7] and $B(U_n)$ is a subcomplex of $B(U_{n+1})$. Hence $M(U_n)$ is a CW-complex and the two-fold suspension

$$S^2M(U_n) = M(U_n \times 1 \times 1)$$

is a subcomplex of $M(U_{n+1})$. Thus

$$\boldsymbol{M}(U) = (0, 0, M(U_1), SM(U_1), M(U_2), \cdots)$$

is a stable object. The track group $\{S^0, \boldsymbol{M}(U)\}_k$ is clearly isomorphic to the stable homotopy group $\pi_{k+2n}(M(U_n))$, with n large.

On the other hand the complexes $B(U_1) \subset B(U_2) \subset \cdots$ have a union $B(U)$ which is again a CW-complex. The isomorphisms

$$\phi\colon H^iB(U_n) \to H^{i+2n}(M(U_n) \bmod o)$$

give rise, in the limit, to an isomorphism

$$\phi\colon H^iB(U) \to H^i\boldsymbol{M}(U).$$

It follows that $H^*\boldsymbol{M}(U)$ has a basis over Z_p consisting of the elements $\phi s(\omega)$, where ω ranges over all partitions.

THEOREM 2. *The cohomology $H^*\boldsymbol{M}(U)$ with coefficient group Z_p is a free module over $A/(Q_0)$, having as basis the elements $\phi s(\lambda)$, where λ ranges over all partitions which contain no integer of the form $p^j - 1$.*

Together with Theorem 1, and the fact that $\boldsymbol{M}(U)$ has no odd dimensional cohomology, this clearly implies the following.

THEOREM 3. *The groups $\{S^0, \boldsymbol{M}(U)\}_m$ have no torsion.*

The full structure of these stable homotopy groups can now be determined, using the fact that the stable Hurewicz homomorphism

$$\{S^0, \mathbf{Y}\}_m \to H_m(\mathbf{Y}; Z)$$

is a $\mathcal{C}$-isomorphism, where $\mathcal{C}$ denotes the class of finite groups. (See Serre [12] for definitions. This particular assertion is not in Serre's paper, but is well known.)

COROLLARY. *The group* $\{S^0, \mathbf{M}(U)\}_m = \pi_m \mathbf{M}(U)$ *is zero for* m *odd, and is free abelian for* $m = 2n$, *the number of generators being equal to the number of partitions of* n.

The proof of Theorem 2 will be based on a peculiar partial ordering of partitions, due to Thom. Given a sequence $R = (r_1, r_2, \cdots)$, define ω_R as the partition of $\sum r_j(p^j - 1)$ consisting of r_j copies of $p^j - 1$ for each $j \geqq 1$. Thus every partition ω can be written uniquely in the form $\lambda\omega_R$, where $\lambda = h_1 \cdots h_l$ contains no integer of the form $p^j - 1$. Let l denote the length of λ and let $\Sigma = h_1 + \cdots + h_l$ denote the sum of the integers in λ. Similarly, given a second partition ω', define l' and Σ'.

Definition. ω' *is less than* ω if $l' < l$, or if $l' = l$ and $\Sigma' > \Sigma$. (Note that integers of the form $p^j - 1$ are completely ignored in this definition.)

LEMMA 6. *The cohomology operation* $\mathcal{P}^R$ *carries* $\phi s(\lambda) \in H^*\mathbf{M}(U)$ *into* $\phi s(\lambda\omega_R)$ *plus a linear combination of elements* $\phi s(\omega')$ *with* ω' *less than* $\lambda\omega_R$.

Proof. It is clearly sufficient to prove the corresponding assertion for $H^*M(U_n)$, where n is large (say $n \geqq l + r_1 + r_2 + \cdots$), but finite. Consider the cross-section

$$f \colon B(U_n) \to E, \partial E$$

of the $2n$-disk bundle, determined by the center points of the disks. The induced cohomology homomorphism f^* carries the fundamental cohomology class $u \in H^{2n}(E \bmod \partial E)$ into the characteristic class

$$c_n = t_1 \cdots t_n = s(1 \cdots 1) \in H^{2n}B(U_n).$$

(See Thom [14], Borel and Serre [4].) Hence f^* carries the general element $\phi(a) = \pi^*(a)u \in H^{i+2n}(E \bmod \partial E)$ into the cup product $ac_n \in H^{i+2n}B(U_n)$. But the correspondence $a \to ac_n$ is a monomorphism; hence f^* is a monomorphism. Thus in order to prove Lemma 6 it is sufficient to prove the following.

Assertion. *$\mathcal{P}^R(s(\lambda)c_n)$ is equal to $s(\lambda\omega_R)c_n$ plus a linear combination of elements $s(\omega')c_n$ with ω' less than $\lambda\omega_R$.*

Consider a typical monomial $t_1^{a_1}\cdots t_n^{a_n}$ of the sum $s(\lambda)c_n$. Here l of the integers $a_1,\cdots,a_n$ are equal to the integers $1+h_1,\cdots,1+h_l$ in some order; while the remaining $n-l$ integers a_i are equal to 1. According to Lemma 3 we have

$$\mathcal{P}^R(t_1^{a_1}\cdots t_n^{a_n})=\sum_{R_1+\cdots+R_n=R}(\mathcal{P}^{R_1}t_1^{a_1})\cdots(\mathcal{P}^{R_n}t_n^{a_n}).$$

This formula is valid even for the case $p=2$, since $B(U_n)$ has no odd dimensional cohomology. (See Lemma 3′.) The expression $\mathcal{P}^{R_i}t_i^{a_i}$ is equal to some constant k_i times $t_i^{b_i}$, where $b_i\geqq a_i$. The case $b_i=a_i$ can occur only if $R_i=0$.

Each such monomial $(k_1\cdots k_n)t_1^{b_1}\cdots t_n^{b_n}$ contributes to a symmetric polynomials $s(\omega')c_n$, where ω' denotes the partition obtained from the sequence $b_1-1,\cdots,b_n-1$ by deleting zero. We wish to choose $R_1,\cdots,R_n$ so that this partition ω' is as "large" as possible, in the sense of the partial ordering. The first requirement is that as few as possible of the integers b_i-1 should be of the form p^j-1. But if $a_i=1$, and if the constant k_i is non-zero, then $\mathcal{P}^{R_i}t_i^{a_i}$ is necessarily of the form $t_i^{p^j}$. (See Lemma 3.) Thus the best we can do is to *choose $R_1,\cdots,R_n$ so that b_i is a power of p only if $a_i=1$.*

The second requirement in order to make ω' "large" is that the sum of all b_i-1 for which b_i is not a power of p should be as small as possible. Evidently, the best we can do in this direction is to *choose $R_i=0$ whenever $a_i>1$*; so that b_i will be equal to a_i whenver $a_i>1$.

Now consider the sum of all terms $(\mathcal{P}^{R_1}t_1^{a_1})\cdots(\mathcal{P}^{R_n}t_n^{a_n})$ for which this last condition (that R_i must be equal to zero whenever $a_i>1$) is satisfied. Each such term has the form $t_1^{b_1}\cdots t_n^{b_n}$, where l of the integers $b_1,\cdots,b_n$ are equal to $1+h_1,\cdots,1+h_l$ in some permutation; and the remaining $n-l$ integers b_i are powers of p. Recall that $\mathcal{P}^{R_i}t_i$ is equal to $t_i^{p^j}$ if $R_i=\Delta_j$ and is zero otherwise. Hence the relation $R_1+\cdots+R_n=R=(r_1,r_2,\cdots)$ implies that a given power p^j, $j\geqq 1$, must occur exactly r_j times in the sequence $b_1,\cdots,b_n$. The integer 1 must therefore occur $n-l-r_1-r_2-\cdots$ times in the sequence $b_1,\cdots,b_n$. Taking the sum of all monomials $t_1^{b_1}\cdots t_n^{b_n}$ which satisfy these conditions, we obtain exactly the polynomial $s(\lambda\omega_R)c_n$. This completes the proof of Lemma 6.

Proof of Theorem 2. The equations

$$\mathcal{P}^R\phi s(\lambda)=\phi s(\lambda\omega_R)+\sum(\text{constant})\phi s(\lambda'\omega_{R'}),$$

with all λ' less than λ, can be solved inductively, giving rise to equations:

$$\phi s(\lambda\omega_R) = \mathcal{P}^R\phi s(\lambda) + \sum(\text{constant})\mathcal{P}^{R'}\phi s(\lambda'),$$

with all λ' less than λ. (Only a finite number of terms are involved, since $H^*\boldsymbol{M}(U)$ is finitely generated in each dimension.) But the elements $\phi s(\lambda\omega_R)$ are known to form a Z_p-basis for $H^*\boldsymbol{M}(U)$. Therefore the elements $\mathcal{P}^R\phi s(\lambda)$ also form a Z_p-basis for $H^*\boldsymbol{M}(U)$. Since $\{\mathcal{P}^R\}$ is a basis for the vector space $A/(Q_0)$ over Z_p, this implies that the elements $\phi s(\lambda)$ form an $A/(Q_0)$-basis for $H^*\boldsymbol{M}(U)$. This completes the proof of Theorem 2, and hence Theorem 3.

4. Cohomology computations for $\boldsymbol{B(SO_{2n})}$ and $\boldsymbol{M(SO)}$. Consider the torus $T_n \subset U_n \subset SO_{2n}$, and the corresponding homomorphism

$$H^*(B(SO_{2n});Z_p) \to H^*(B(T_n);Z_p).$$

According to Borel and Serre [4], if p is odd, then the first algebra may be identified with the subalgebra of the second consisting of all polynomials $a + t_1 \cdots t_n b$, where a and b are symmetric polynomials in the elements $t_1^2, \cdots, t_n^2$. Thus a basis for $H^*(B(SO_{2n});Z_p)$ over Z_p is given by the elements $s(\omega)$ and $s(\omega)t_1 \cdots t_n$, where $\omega = i_1 \cdots i_k$, $k \leqq n$, is a partition into even integers. Letting n tend to infinity, *a Z_p-basis for $H^*(B(SO);Z_p)$ is given by the elements $s(\omega)$, where ω ranges over all partitions into even integers.*

Carrying out an argument completely analogous to that in Section 3, we construct a stable object

$$\boldsymbol{M}(SO) = (o, M(SO_1), M(SO_2), \cdots),$$

and prove the following.

THEOREM 4. *Let p be an odd prime, and let $\lambda = h_1 \cdots h_l$ range over all partitions into integers h_i which are even and not of the form $p^j - 1$. Then $H^*(\boldsymbol{M}(SO);Z_p)$ is the free $A/(Q_0)$-module having as basis the elements $\phi s(\lambda)$.*

Together with Theorem 1 this proves the following

THEOREM 5. *The cobordism groups $\Omega^i = \pi_i(\boldsymbol{M}(SO))$ contain no odd torsion.*

C. T. C. Wall has recently proved that an element in the 2-torsion subgroup of Ω^i is completely determined by its Stiefel-Whitney numbers. Together with Theorem 5, this proves the following conjecture of Thom.

COROLLARY 1. *If the Stiefel-Whitney numbers and the Pontrjagin numbers of a compact, oriented, differentiable manifold V^i are all zero, then V^i is a boundary.*

As special cases:

COROLLARY 2. *Suppose that V^i can be imbedded in euclidean space so as to have trivial normal bundle. Then V^i is a boundary.*

The proof is clear.

COROLLARY 3. *Suppose that $H_*(V^i; Z_2)$ is isomorphic to $H_*(S^i; Z_2)$. Then V^i is a boundary.*

Proof. The Stiefel-Whitney number $w_i[V^i]$ is equal to the Euler characteristic reduced modulo 2; hence is zero. If $i = 4n$, then the Pontrjagin number $p_n[V^i]$ is zero by the index theorem (Hirzebruch [8]). Since the other characteristic numbers of V^i are trivially zero, it follows that V^i is a boundary.

Concluding Remarks. There are other homotopy groups which may be accessible, using the Adams spectral sequence. For example, the symplectic groups $Sp(n) \subset SO_{4n}$ give rise to a stable object

$$\boldsymbol{M}(Sp) = (o, o, o, o, M(Sp(1)), SM(Sp(1)), S^2M(Sp(1)), S^3M(Sp(1)), M(Sp(2)), \cdots).$$

Assertion. The groups $\pi_i\boldsymbol{M}(Sp)$ have no odd torsion.

This can be proved directly from the spectral sequence; or can be derived from Theorem 5, using the natural map $\boldsymbol{M}(Sp) \to \boldsymbol{M}(SO)$.

Problem. Can one compute the spectral sequence for $\pi_*\boldsymbol{M}(Sp)$ corresponding to the prime $p = 2$?

Similarly, the representations $Spin(n) \to SO_n$ give rise to a stable object.

$$\boldsymbol{M}(Spin) = (o, M(Spin(1)), M(Spin(2)), \cdots).$$

Again there is no odd torsion; but the case $p = 2$ seems difficult. As a final question, consider the stable object $\boldsymbol{M}(SU)$ corresponding to the special unitary group.

Problem. What can be said about $\pi_*\boldsymbol{M}(SU)$?

REFERENCES.

[1] J. F. Adams, "On the structure and applications of the Steenrod algebra," *Commentarii Mathematici Helvetici*, vol. 32 (1958), pp. 180-214.

[2] ———, "Sur la théorie de l'homotopie stable," *Bulletin de la Société Mathématique de France*, vol. 87 (1959), pp. 277-280.

[3] J. Adem, "The relations on Steenrod powers of cohomology classes," *Algebraic geometry and topology*, Princeton, 1957, pp. 191-238.

[4] A. Borel and J.-P. Serre, "Groupes de Lie et puissances réduites de Steenrod," *American Journal of Mathematics*, vol. 75 (1953), pp. 409-448.

[5] H. Cartan, "Sur les groupes d'Eilenberg-MacLane II," *Proceedings of the National Academy of Sciences U. S. A.*, vol. 40 (1954), pp. 704-707.

[6] ———, "Sur l'iteration des opérations de Steenrod," *Commentarii Mathematici Helvetici*, vol. 29 (1955), pp. 40-58.

[7] C. Ehresmann, "Sur la topologie de certains espaces homogenes," *Annals of Mathematics*, vol. 35 (1934), pp. 396-443.

[8] F. Hirzebruch, *Neue topologische Methoden in der algebraischen Geometrie*, Springer, 1956.

[9] E. L. Lima, "Stable Postnikov invariants," *American Mathematical Society Notices*, vol. 5 (1958), abstract number 544-27, p. 215.

[10] J. Milnor, "The Steenrod algebra and its dual," *Annals of Mathematics*, vol. 67 (1958), pp. 150-171.

[11] F. Peterson, "Generalized cohomotopy groups," *American Journal of Mathematics*, vol. 78 (1956), pp. 259-281.

[12] J.-P. Serre, "Groupes d'homotopie et classes de groupes abéliens," *Annals of Mathematics*, vol. 58 (1953), pp. 258-294.

[13] E. Spanier, "Function spaces and duality," *Annals of Mathematics*, vol. 70 (1959), pp. 338-378.

[14] R. Thom, "Espaces fibrés en sphères et carrés de Steenrod," *Annales Scientifiques de l'Ecole Normale superieure*, vol. 69 (1952), pp. 109-181.

[15] ———, "Quelques propriétés globales de variétés différentiables," *Commentarii Mathematici Helvetici*, vol. 28 (1954), pp. 17-86.

SOLUTION OF SOME PROBLEMS OF DIVISION.*

Part IV. Invertible and Elliptic Operators.[1]

By L. EHRENPREIS.

1. Introduction. Let V be a topological vector space and $L\colon V \to V$ a continuous linear map. The study of the linear equation

$$(1) \qquad Lv = w \qquad (v, w \in V)$$

leads us to three very natural questions:

Problem A. What is the image of L, that is, for what w does (1) have a solution?

Problem B. Let w have some additional properties, that is, let w belong to some subspace W of V; what additional properties does v have?

Problem C. What is the lack of uniqueness of (1)? That is, describe completely the kernel of L, the set of v which satisfy $Lv = 0$.

In this series we study mainly the following case: V is a space of distributions or functions, and L is a convolution $Lv = a * v$, where a is a certain distribution such that $v \to a * v$ is a continuous map of V into V. For many spaces V, which can be describel roughly by saying that the Fourier transform of the dual V' of V is a space of entire functions which is described by growth conditions at infinity, these growth conditions depending only on the distance from the origin, it was shown in Part III (see [7]) that L is always onto. On the other hand, for many interesting spaces, e. g., the space $\mathcal{E}$ of indefinitely differentiable functions on $R = R^n$ = Euclidean n space or the $\mathcal{D}'$ of distributions on R (see [24]) there exist continuous convolution maps which are not onto (an example is convolution by an indefinitely differentiable function of compact carrier). Thus, problem A takes on a non-trivial form for these spaces. We shall give first a partial solution to problem A by solving completely

* Received October 5, 1959.

[1] Work partially supported by National Science Foundation Grant NSF 5-G1010.

Problem A′. Determine those L for which L' is onto.

Definition. If $L\colon V \to V$ is onto, we call L *invertible.* If $Lv = \alpha * v$, we say α is invertible if L is.

Thus, we shall determine all invertible elements of $\mathcal{E}'$ considered as convolution operators of $\mathfrak{D}'$ into $\mathfrak{D}'$, or $\mathcal{E}$ into $\mathcal{E}'$.

We denote by $\boldsymbol{E}'$ the Fourier transform of $\mathcal{E}'$; thus $\boldsymbol{E}'$ is the space of all entire functions of exponential type on $C = C^n$ which are of polynomial increase on $R = R^n$ which is the real part of C (see [24], [11]).

Definition. A function $J \in \boldsymbol{E}'$ is called *slowly decreasing* if there exists a positive number a such that for each point $x \in R$ we can find a point $y \in R$ with

$$|x - y| \leqq a \log(1 + |x|) \tag{2}$$

$$|J(y)| \geqq (a + |y|)^{-a}. \tag{3}$$

Then our first main result is

THEOREM I. *$S \in \mathcal{E}'$ is invertible for $\mathfrak{D}'$ (or for $\mathcal{E}$) if and only if the Fourier transform of S is slowly decreasing. In order for S to be invertible it is sufficient to be able to solve the equation $S * T = S'$ for $T \in D'$, where S' is a fixed invertible distribution in E'.*

Since $\delta =$ Dirac's measure is clearly invertible, Theorem I contains as a special case a conjecture of L. Schwartz (see [24]).[2]

Theorem I was announced in a Proceedings note [9].

In case S is not invertible, we could, of course, again ask the question (Problem A) as to what is $S * \mathfrak{D}'$ of $S * \mathcal{E}$. In Part II (see [7] p. 692) we stated the conjecture that if $S \in \mathfrak{D}$ then $S * \mathfrak{D}' \neq \mathcal{E}$. We shall show that, in fact, much more is true (see Theorem 2.5 below) : If S is not invertible, then it is not even true that $\mathfrak{D} \subset S * \mathfrak{D}'$.

We can also give a more complete answer to Problem A of determining the image $S * \mathfrak{D}'$ (or $S * \mathcal{E}$) for any $S \in \mathcal{E}'$. A trivial necessary and sufficient condition for $T \in \mathfrak{D}'$ to be in $S * \mathfrak{D}'$ is that, if $f_\alpha \in \mathfrak{D}$, $S * f_\alpha \to 0$ in $\mathfrak{D}$, then $T \cdot f_\alpha \to 0$, that is, T is continuous on the topology τ defined on $\mathfrak{D}$ as follows: N is a neighborhood of zero in τ if we can find a neighborhood N' of zero in $\mathfrak{D}$ so that N consists of all $f \in \mathfrak{D}$ with $S * f \in N'$. Now, a good description

[2] It is incorrectly stated by Schwartz in [24] that the result for $n = 1$ can be proved by the methods of the theory of mean periodic functions. Professor Schwartz has kindly pointed out to me that his proof only shows that if $S * T = \delta$ has a solution $T \in \mathfrak{D}'_F$, then $S * \mathfrak{D}'_F = \mathfrak{D}'_F$ (see below). This result was extended to $n > 1$ by Malgrange [21].

of $S * \mathfrak{D}'$ would be obtained if we give a "good" description of τ. This seems to be extremely difficult and we shall content ourselves with giving a partial solution to this problem (Section 2).

In addition, we shall give in Section 2 several other necessary and sufficient conditions for invertibility of S. One such is (see Theorem 2.6 below): For any entire function G, $JG \in \boldsymbol{D}$ implies $G \in \boldsymbol{D}$. (Here J is the Fourier transform of S and $\boldsymbol{D}$ is the Fourier transform of the space $\boldsymbol{D}$ of L. Schwartz [24].)

We shall prove the analog of Theorem I for the space $\mathfrak{D}'_F$ of distributions of finite order.

It might be suspected that for $S \in \mathcal{E}'$, even if S is not invertible for $\mathfrak{D}'$, we should be able to find a larger space $\mathfrak{D}'_M$ which is the dual of a space of Carleman non quasi-analytic functions (see e.g. [14]) such that for some $T \in \mathfrak{D}'_M$ we can solve the equation $S * T = \delta$. However, I shall produce an S for which there cannot exist a $T \in \mathfrak{D}'_M$ for any Carlemann non quasi-analytic class (see Theorem 6.2 below). I should only remark that the existence of this example came as a very great surprise to me personally.

Section 3 is devoted towards proving that in case S is invertible (for $\mathfrak{D}'$ or $\mathcal{E}'$) then L. Schwartz' mean periodic expansion (see [27]) for a solution f of $S * f = 0$ in terms of the exponential polynomial solutions can be greatly simplified.

The next question which we shall discuss (Section 4) is: Let $T \in \mathfrak{D}'$ have the property that for any $S \in \mathcal{E}'$ the equation $S * U = T$ has a solution $U \in \mathfrak{D}'$; what can be said about T? That is what is the intersection of all $S * \mathfrak{D}'$ for $S \in \mathcal{E}'$? We shall prove

THEOREM II. *For $T \in D'$, a necessary and sufficient condition that for each $S \in \mathcal{E}'$ the equation $S * U = T$ has a solution $U \in \mathfrak{D}'$ is that T be real analytic. In fact, $\bigcap_{S \in \mathcal{E}'} S * \mathfrak{D}' = \bigcap_{S \in \mathcal{E}'} S * \mathcal{E} =$ real analytic functions. In case T is analytic in a strip in S around R, then for any $S \in \mathcal{E}'$ there exists a U which is analytic in a strip in C around R such that $S * U = T$.*

The proof of Thetorem II depends on the Denjoy-Carleman Theorem for quasi-analytic functions (see [22]).

In Section 5 we discuss a special case of Problem B: We say that the distribution T is C^∞ in x_1 if we can find a sequence of positive numbers a_j so that $\{a_j(\partial^j/\partial x_1{}^j)T\}$ is a bounded set in $\mathfrak{D}'$, or, what can be seen to be the same thing, that T belongs to the topological tensor product (see [19]) of $\mathcal{E}(x_1)$ with $\mathfrak{D}'(x_2, \cdots, x_n)$. (A proof of this assertion can be obtained by

use of methods of [12].) Then we want to find all $S \in \mathcal{E}'$ so that if $S * U$ is C^∞ in x_1, then U is C^∞ in x_1. If S has this property, then we say that S is C^∞ elliptic in x_1. Our main result for this problem is

THEOREM III. *Let $S \in \mathcal{E}'$ be invertible for $\mathcal{D}'$. S is C^∞ elliptic in x_1 if and only if for each $r \geqq 0$ we can find a $b_r > 0$ with the property that*

$$|\mathfrak{J}z| \geqq r \log(1 + |z_1|) \tag{1.1}$$

whenever $z \in V$ and

$$|z_1{}^r| \geqq b_r(1 + |z|). \tag{1.2}$$

It should be remarked that conditions (1.1) and (1.2) could be condensed to the single condition:

$$|z_1|^r \exp(-d\,|\mathfrak{J}z|) \leqq d_r(1 + |z|)^d. \tag{1.3}$$

A similar remark applies to all the other cases of ellipticity discussed below.

We also show by an example (Example 4 of Section 5 below) that if S is not invertible for $\mathcal{D}'$, then conditions (1.1) and (1.2) do not suffice even to guarantee that all distribution solutions of $S * T = 0$ be C^∞ in x_1.

We call the distribution T *entire* in x_1 if for any $a > 0$ the set $\{(a^j/j!)(\partial^j T/\partial x_1{}^j)\}$ is bounded in $\mathcal{D}'$ or, what is the same thing, if T belongs to the topological tensor product of the space of entire functions in x_1 with the space $\mathcal{D}'(x_2, \cdots, x_n)$. We say that $S \in E'$ is entire elliptic in x_1 if whenever $S * T$ is entire in x_1, then T is entire in x_1. We prove the analog of Theorem III for entire ellipticity.

We could define similarly classes between C^∞ ellipticity and entire ellipticity, but we shall not do this since the methods of the present paper apply without essential modifications.

We prove also that if S is C^∞ elliptic in all variables, then it is necessarily invertible.

In the above we have used the space $\mathcal{D}'$ to define ellipticity. We could make a similar definition for the space $\mathcal{D}'_F$ (see [5]) of distributions of finite order; we call this ellipticity *weak* ellipticity (thus, weak C^∞ elliptic; weak elliptic, etc.). Now there is no difference between weak entire ellipticity and entire ellipticity, and if S is a differential operator, then weak C^∞ ellipticity and C^∞ ellipticity are the same. However, if $S \cdot f = df(0)/dx + f(1)$, then S is weakly C^∞ elliptic but S is not C^∞ elliptic.

Finally, we show certain special properties of elliptic operators. For example, if S is entire elliptic in x_1, then, in x_1, S is the composition of a

translation with a differentiation. Thus, if S is entire elliptic in all variables, then it is the composition of a classical elliptic differental operator with a translation. If S is a differential difference operator in x_1 which is C^∞ elliptic in x_1, then in x_1, S is the composition of a translation with a differentiation. Thus, if S is a differential-difference operator in all variables which is C^∞ elliptic in all variables, then S is the composition of a translation with a partial differential operator which is C^∞ elliptic in all variables. The above example $S\cdot f = df(0)/dx + f(1)$ shows that the analog of the last proposition is false for weak C^∞ ellipticity.

Suppose that $T \in \mathfrak{D}'$ is C^∞ in $x_1, \cdots, x_r$ $(r < n)$. Then we prove easily (see Proposition 5.1) that for h an indefinitely differentiable of compact support in $x_{r+1}, \cdots, x_n$ we have $T * h \in \mathcal{E}$. (Here $T * h$ is the convolution of T with the direct product of h with $\delta(x_1, \cdots, x_r)$.) However, the converse is not true, e.g., for such an h we have $\delta(x_1 - x_2) * h(x_2) = h(x_1 - x_2) \in \mathcal{E}$, while $\delta(x_1 - x_2)$ is not C^∞ in x_1. If T has the property that $T * h \in \mathcal{E}$ for all such h, then we can prove easily by means of the closed graph theorem that $h \to (T * h)(0)$ is a distribution on these h which we call the restriction of T to the plane $x_1 = x_2 = \cdots = x_r = 0$.

This leads to the following concept: Let us partition the variables $x = (x_1, x_2, \cdots, x_n)$ into three sets, say $x = (x', x'', x''')$. Then we say that $T \in \mathfrak{D}'$ is C^∞ *in* x' *relative to* x''' if for any $h \in \mathfrak{D}(x''')$, we have $T * h$ is C^∞ in the variables x'. We say that $S \in \mathcal{E}'$ is C^∞ *elliptic in* x' *relative to* x''' if whenever $T \in \mathfrak{D}'$ is C^∞ in x' relative to x''' and $W \in \mathfrak{D}'$ satisfies $S * W = T$, then W is also C^∞ in x' relative to x'''. For simplicity of notation shall consider only the case when the variables x'' are absent, but there is no difficulty in extending all our results to the general case.

In Section 5 we describe all invertible $S \in \mathcal{E}'$ which are C^∞ elliptic in x' relative to x'''. In particular, if S is a differential operator in x_1 with leading coefficient 1, then S is C^∞ elliptic in x_1 relative to $(x_2, \cdots, x_n)$. Thus, a distribution solution of any linear constant coefficient partial differential equation has a restriction to any non-characteristic hypersurface.

We can define also similar concepts for entire ellipticity and weak ellipticity. The classes of $S \in \mathcal{E}'$ which satisfy these conditions are described in Section 5.

The results of previous authors on the problem of ellipticity deal with the case where S is a linear partial differential operator, and the only kind of ellipticity considered is simultaneous ellipticity in all variables.[3] In this

[3] I have been just informed by L. Gårding that he and B. Malgrange have characterized all linear constant coefficient partial differential operators which are C^∞ elliptic in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$.

case, the analog of Theorem III for entire ellipticity was found by Petrowski [27], while the analog for C^∞ ellipticity was obtained by Hörmander [20]. We should like to mention that the results of Theorem III were announced in the Proceedings of the National Academy of Sciences [10] and they were obtained independently of those of Hörmander (though Hörmander's results appeared slightly before mine).

In case S is a linear constant coefficient partial differential operator, then a complete solution to Problem C will be given in a future publication (see [15], [16]).

The problem of extending the results of this paper to continuous linear transformations $L: \mathfrak{D}' \to \mathfrak{D}'$ which are not of the form $L(T) = S * T$ for some $S \in \mathcal{E}'$ is undoubtedly very difficult.

The paper concludes with a list of unsolved problems and some general remarks.

The notations in this paper will be the same as those used in Parts I, II, and III.

I should like to thank my friend Dr. D. J. Phlotzelphlip for several useful discussions. I should also like to thank Professor Beurling from whom I have learned much.

Added in proof. Since this paper was written, some of the questions posed have been answered. I have inserted an appendix at the end to give some indications of the progress.

2. Invertible operators for $\mathfrak{D}'$, $\mathcal{E}'$, $\mathfrak{D}'_F$. In this section we shall first find all invertible operators for $\mathfrak{D}'$. In a previous paper [7] we proved that all differential-difference operators were invertible for $\mathfrak{D}'$. The proof consisted essentially of two parts: (a) Describe explicitly the topology of the Fourier transform $\boldsymbol{D}'$ of $\mathfrak{D}$. (b) Use the fact that exponential polynomials (the Fourier transform of differential-difference operators) do not tend to zero too fast at infinity "too often," and use the minimum modulus theorem to take care of the points where the exponential polynomials are small. We want to give a simplified abstract treatment of (a) and (b).

Let then A be a topological vector space of entire functions on C. We assume that the topology of A can be described as follows: There exist continuous positive functions $\{H\}$ on C so that a fundamental system of neighborhoods of zero in A consists of the sets N_H comprising all functions $f \in A$ such that $|f(z)| \leqq H(z)$ for all $z \in C$.

This is part (a) described above; now to part (b): Let $z_0 \in C$ and let α

be a subset of C; we say that α *surrounds* z_0 if for every entire function g

$$|g(z_0)| \leqq \max_{z\in\alpha} |g(z)|. \tag{2.1}$$

Now, let J be an entire function for which $f \to Jf$ defines a continuous map of $A \to A$. We want to know when $Jf \to f$ is a continuous map of $JA \to A$. (JA is the space of Jf, $f \in A$ with the topology induced from A.) Suppose each point $z_0 \in C$ can be surrounded by a set α on which J is large, say $|J(z)| \geqq b(z_0)$ for all $z \in \alpha$. Then how large does $b(z_0)$ have to be in order to guarantee that $Jf \to f$ is continuous? Let H be one of the functions above used to define the topology of A; when can we find another such function H' so that the conditions $|J(z)f(z)| \leqq H'(z)$ should imply $|f(z)| \leqq H(z)$? We know from the above that if $|J(z_0)f(z_0)| \leqq H'(z_0)$, then on α we have $|f(z)| \leqq H'(z)[b(z_0)]^{-1}$. Thus, since α surrounds z_0 and f is entire, it follows that

$$|f(z_0)| \leqq [b(z_0)]^{-1} \max_{z\in\alpha} H'(z). \tag{2.2}$$

Putting the above together we have

LEMMA 2.1. *With the above notation, suppose for every H we can find an H' so that for all $z_0 \in C$,*

$$[b(z_0)]^{-1} \max_{z\in\alpha} H'(z) \leqq H(z_0), \tag{2.3}$$

then $Jf \to f$ is a continuous linear map of $JA \to A$.

Finally, there remains the problem of constructing the sets α. We shall construct them by means of the minimum modulus theorem. We suppose now that J is an entire function of exponential type B which is bounded on R (the case of an arbitrary $J \in \boldsymbol{E}'$ is easily reduced to this). Suppose that for each $x \in R$ we can find a $y \in R$ so that $|x-y| \leqq p(x)$ and $J(y) \geqq q(x)$. Let $y_2, \cdots, y_n$ be fixed and draw about the point y_1 in the complex plane a circle $\beta(y_2, \cdots, y_n)$ of center y_1 and radius between $p(x)$ and $2p(x)$ on which

$$|J(w_1, y_2, \cdots, y_n)| \geqq M\,|J(y_1, y_2, \cdots, y_n)|^d \exp(-dBp(x)). \tag{2.3}$$

Here d is a positive integer and M depends only on J. The existence of such a circle is guaranteed by the minimum modulus theorem (see [8]). Moreover, it is clear that if $x_2 = y_2, \cdots, x_n = y_n$, then x is surrounded by this circle.

If it is not true that x is surrounded by β, then we just iterate the above process. For $w_1 \in \beta, y_3, \cdots, y_n$ fixed, we can draw in the complex plane a

circle $\beta(w_1, y_3, \cdots, y_n)$ of center y_2 and radius between $p(x)$ and $2p(x)$ on which

$$(2.4) \qquad |J(w_1, w_2, y_3, \cdots, y_n)| \geqq M_1 |J(y_1, y_2, \cdots, y_n)|^d \exp(-2dBp(x)).$$

The existence of such a circle is again guaranteed by the minimum modulus theorem. Suppose for simplicity that $y_3 = x_3, y_4 = x_4, \cdots, y_n = x_n$. Then I claim we could choose

$$(2.5) \qquad \alpha = \{(w_1, w_2, y_3, \cdots, y_n)\}_{w_2 \in \beta(w_1, y_3, \cdots, y_n)}.$$

For, given any entire function $g(w)$, we have for any $w_1 \in \beta(y_2, y_3, \cdots, y_n)$,

$$(2.6) \qquad |g(w_1, x_2, y_3, \cdots, y_n)| \leqq \max_{w_2 \in \beta(w_1, w_2, y_3, \cdots, y_n)} |g(w_1, w_2, y_3, \cdots, y_n)|$$

by the maximum modulus theorem. Hence, again by the maximum modulus theorem, we have

$$(2.7) \qquad |g(x_1, x_2, y_3, \cdots, y_n)| \leqq \max_{w_1 \in \beta(y_2, y_3, \cdots, y_n)} |g(w_1, x_2, y_3, \cdots, y_n)|.$$

Thus, our assertion is proven.

Finally, if it is not true that $y_3 = x_3, y_4 = x_4, \cdots, y_n = y_n$, we just continue the above process.

In case $y \in R$ but $x \notin R$, the same considerations apply. We have thus proven

LEMMA 2.2. *In the above notation, given any $z_0 \in R$, we can surround z_0 by a set α on which*

$$(2.8) \qquad |J(z)| \geqq M[q(x)]^d \exp(-dBp(x)),$$

where d and M are constants depending only on J. Morevoer, we have

$$(2.9) \qquad \max_{z \in \alpha} |z - z_0| \leqq Mp(z_0).$$

Now we are ready to apply the above to the space $\boldsymbol{D}$. For this purpose, we have to describe the topology of $\boldsymbol{D}$ in a manner similar to the space A above. This is

THEOREM 2.1. *Let $H(z)$ be any continuous positive function on C such that if h is any continuous function on C for which we can find an $a > 0$ so that*

$$(2.10) \qquad h(z) = O(\exp(a|\mathfrak{J}(z)|)/(1 + |z|^m)$$

for all m, then also $h(z) = O(H(z))$. Call N_H the set of $F \in \boldsymbol{D}$ for which $|F(z)| \leqq H(z)$ for all $z \in C$. Then the sets N_H form a fundamental system of neighborhoods of zero in $\boldsymbol{D}$.

Proof. First we show that the sets N_H are neighborhoods of zero in $\boldsymbol{D}$. Now, each N_H is convex and clearly for any $F \in \boldsymbol{D}$ we can find a $b > 0$ so that $bF \in N_H$. Let B be a bounded set in $\boldsymbol{D}$; set

$$h(z) = \max_{F \in B} |F(z)| \tag{2.11}$$

for all $z \in C$. Then, by the explicit description of the bounded sets of $\boldsymbol{D}$ (see [5], [11]), we see that h satisfies (2.10). Thus, we can find a $b' > 0$ so that $b'B \subset N_H$. Since $\boldsymbol{D}$ is bornologic (see [4]), it follows that N_H is a neighborhood of zero in $\boldsymbol{D}$.

We now have to show that the neighborhoods N_H are fundamental in $\boldsymbol{D}$. But this is an immediate consequence of Theorem 1 of [7] which shows that sets N_H can be used to describe the topouogy of $\boldsymbol{D}$.

We are now in a position to prove our first main theorem (from which Theorem I follows immediately):

THEOREM 2.2. *For $S \in \boldsymbol{E}'$ the following properties are equivalent:*

(a) *$\mathfrak{F}(S)$ is slowing decreasing.*

(b) *$S * f \to f$ is a continuous linear map of $S * \mathfrak{D} \to \mathfrak{D}$.*

(c) *$S * \mathfrak{D}' = \mathfrak{D}'$.*

(d) *There exists an invertible $S' \in \mathcal{E}'$ such that $S * T = S'$ has a solution $T \in \mathfrak{D}'$.*

(e) *$S * U \to U$ is a semi-continuous linear map of $S * \mathcal{E}' \to \mathcal{E}'$.* (That is, the map sends bounded sets into bounded sets.)

Proof. We shall prove Theorem 2.2 by the usual chain of implications: (a) implies (b), (b) implies (c), etc. We shall prove the simpler parts first.

(b) *implies* (c): This is an immediate consequence of the Hahn-Banach theorem.

(c) *implies* (d): This is a triviality.

(d) *implies* (e): Let $S * B$ be bounded. Then it follows that $S * B * T$ is bounded in $\mathfrak{D}'$, that is, $S' * B$ is bounded in $\mathfrak{D}'$ by the associativity and commutativity of convolution (see [24]). But all the $S * U$ for $U \in B$ have their carriers in a fixed compact set; hence, all $U \in B$ have their carrier in a fixed compact set by the theorem on addition of carriers (see [24]). Thus, $S' * B$ is a set which is bounded in $\mathfrak{D}'$ and all the distributions in $S' * B$ have their carriers in a fixed compact set. Thus (see [24]) $S' * B$ is bounded in $\mathcal{E}'$.

Now we can find a $T' \in \mathfrak{D}'$ such that $S' * T' = \delta$. Repeating the above argument we find that $\delta * B = B$ is bounded in $\mathcal{E}'$. This completes the proof that (d) implies (e).

We now have to prove the two difficult parts of our theorem:

Proof that (a) *implies* (b). We shall use the Lemmas 2.1 and 2.2 and Theorem 2.1. We assume first that $J = \mathfrak{J}(S)$ is bounded on R; we shall later dispense with this assumption. Let H be any function as described in Theorem 2.1. For each $z_0 \in C$ we choose an α satisfying the conclusions of Lemma 2.2; here we can take $p(z_0) = |\mathcal{J}(z_0)| + A\log(1+|z_0|)$, and and $q(x) = (A+|x|)^{-A}$. In the notation of Lemma 2.1 we can choose therefore by Lemma 2.2,

$$\begin{aligned} b(z_0) &= M(A+|x|)^{-dA}\exp(-dB(|\mathcal{J}(z_0)| + A\log(1+|z_0|))) \\ &= M(A+|x|)^{-dA}(1+|z\ |)^{-daB}\exp(-dB\,|\mathcal{J}(z_0)|). \end{aligned} \tag{2.12}$$

Since $|z_0 - x| \leqq |\mathcal{J}(z_0)| + A\log(1+|z_0|)$, we have

$$\begin{aligned} A+|x| &\leqq A+|z_0|+|z_0-x| \\ &\leqq A+2|z_0|+A\log(1+|z_0|) \\ &\leqq A'+3|z_0| \end{aligned}$$

for some A'. Thus, we can assume by changing M if necessary, that

$$b(z_0) = M(M+|z_0|)^{-M}\exp(-M\,|\mathcal{J}z_0|). \tag{2.13}$$

Thus, according to Lemma 2.1, we have to find an H' such that

$$\max_{z\in\alpha} H'(z) \leqq M(M+|z_0|)^{-M}\exp(-M\,|\mathcal{J}z_0|)H(z_0). \tag{2.14}$$

It is now clear that if H' exists then, by Lemma 2. 2, (2.9), a good choice would be

$$H'(z) = \min_{|z-z_0|\leqq M|\mathcal{J}z_0|+M\log(1+|z_0|)} M(M+|z_0|)^{-M}\exp(-M\,|\mathcal{J}z_0|)H(z_0). \tag{2.15}$$

We have only to show that this choice of H' satisfies the hypotheses of Theorem 2.1. Let then $h(z)$ be a continuous function for which we can find an $a > 0$ so that for all $m > 0$, (2.10) is satisfied. To show that $h(z)/H'(z)$ is bounded, we have to show that if we define

$$h'(z_0) = M^{-1}(M+|z_0|)^{M}\exp(M\,|\mathcal{J}z_0|) \max_{|z-z_0|\leqq M|\mathcal{J}z_0|+M\log(1+|z_0|)} h(z), \tag{2.16}$$

then $h'(z_0)$ again satisfies (2.10). Thus, we have only to show that

$$h''(z_0) = \max_{|z-z_0|\leqq M|\mathcal{J}z_0|+M\log(1+|z_0|)} h(z) \tag{2.17}$$

satisfies (2.10) whenever h does. To show that

$$h''(z_0)\exp(-M'\,|\mathcal{J}z_0|)(1+|z_0|)^{m}$$

is bounded for a suitable M' is the same as showing that

$$(2.18)\qquad h(z) \max_{|z-z_0|\leqq M|\mathfrak{J}z_0|+M\log(1+|z_0|)} \exp(-M'|\mathfrak{J}z_0|)(1+|z_0|)^m$$

is bounded.

Consider $\min_{|z-z_0|\leqq M|\mathfrak{J}z_0|+M\log(1+|z_0|)} |\mathfrak{J}z_0|$. It is easily seen that if $|\mathfrak{J}z| \geqq 2M\log(1+|z|)$ then this minimum is $\geqq (4M)^{-1}|\mathfrak{J}z|$. Thus, if $|\mathfrak{J}z| \geqq 2M\log(1+|z|)$,

$$\max_{|z-z_0|\leqq M|\mathfrak{J}z_0|+M\log(1+|z_0|)} \exp(-M'|\mathfrak{J}z_0|) \leqq \exp(-M''|\mathfrak{J}z|)$$

for a suitable M''. Moreover, for all z, $\max\exp(-M'|\mathfrak{J}z_0|) \leqq 1$. Thus we can find an M''' so that for all z we have

$$(2.19)\qquad \max\exp(-M'|\mathfrak{J}z_0|) \leqq \exp(-M''|\mathfrak{J}z|)(1+|z|)^{M'''}.$$

Clearly we also have

$$(2.20)\qquad \max_{|z-z_0|\leqq M|\mathfrak{J}z_0|+M\log(1+|z_0|)} \exp(-|\mathfrak{J}z_0|)(1+|z_0|)^m \leqq m'(1+|z|)^{m'}$$

for a suitable m' depending only on m.

Putting (2.20), (2.19) together with (2.18) we have the desired result that $h''(z_0)$ satisfies (2.10) when h does. Thus H' defined by (2.15) satisfies the hypotheses of Theorem 2.1 and the desired implication (a) *implies* (b) is established when J is bounded on R.

In case J is not bounded on R, set $K(z) = \sin z_1 \sin z_2 \cdots \sin z_n / z_1 z_2 \cdots z_n$. Then for l large enough, $J' = K^l J$ is bounded on R. Since the first partial derivatives of J on R are $O(1+|x|^p)$ for some p, we see easily that J' is again slowly decreasing. If $JF \to 0$ in $\boldsymbol{D}$, then so does $J'F$. Hence by the above, $F \to 0$ in $\boldsymbol{D}$. Thus, (a) *implies* (b) is proven in all cases.

(Instead of reducing the case of J unbounded to the case J bounded, we could have argued directly as we did in the bounded case.)

To complete the proof of Theorem 2.2 we have to show (e) *implies* (a): Assume then that J is not slowly decreasing; we have to find a set $B \subset \mathcal{E}'$ so that JB is bounded in $\mathcal{E}'$ but B is not bounded in $\mathcal{E}'$.

For JB to be bounded in $\mathcal{E}'$ all the functions $F \in B$ have to be bounded exponential type (say $\leqq \pi$). If we want B to be unbounded, then we want to make F large where J is small. Since J is small on large sets, such functions F can be constructed which are very large.

Since J is not slowly decreasing we can find a sequence of points $x^j \in R$ with the following properties:

α. $|x^j| > e^{3j}$.

β. On the set $\{y \in R, |y - x^j| \leqq j \log |x^j|\}$, we have $|J(y)| \leqq |y|^j$.

For each point x^j we want to find first an $F'_j \in E'$ so that $F'_j(x^j) \geqq (1/e)|x^j|^j$ but $|JF'_j(x)| \leqq 1$ for all $x \in R$. Moreover, we want to make sure that F'_j is of exponential type $\leqq \pi$.

We shall construct F'_j as $F'_j(z) = F_j(z_1)F_j(z_2) \cdots F_j(z_n)$, where $F_j \in {}_1\boldsymbol{D}$ (space $\boldsymbol{D}$ in one variable). Let us assume first that $n = 1$; the general case will be similar. Then we want $F_j(x^j) \geqq (1/e)|x^j|^j$ and then we want F_j to decrease very rapidly. By the minimum modulus theorem, F_j cannot decrease for a long distance faster than $\exp(-a(\text{distance to } x^j))$ for some $a > 0$. On the other hand, F_j cannot decrease exponentially for too long a distance, for this would contradict known inequalities on $\int_{-\infty}^{\infty} (\log |F_j(x)|/(1 + |x|^2))\,dx$. Thus, we want to construct F_j so that

a. $F_j(x^j) \geqq (1/e)|x^j|^j$.
b. $F_j(x)$ decreases exponentially as long as possible until it reaches a value $\leqq 1$.
c. $F_j(x)$ stays $\leqq 1$ after that point.
d. F_j is of exponential type $\leqq \pi$.

Let us define H_j by

$$H_j(z) = \prod_{k=1}^{\infty} (1 - z^2/j^2k^2)^j = ((j/\pi z)\sin(\pi z/j))^j.$$

Then the following properties are readily verified:

1. H_j is an entire function of exponential type π.
2. $H_j(0) = 1$.
3. $|H_j(x)| \leqq 1$ for x real.
4. $|H_j(x)| \leqq e^{-j}$ for $x \in R$, $|x| \geqq j$.

Then we set

$$F_j(z) = e^k H_k(z - x^j), \tag{2.21}$$

where k is the greatest integer in $j \log |x^j|$. Then $F_j(z)$ has the following properties:

1′. F_j is an entire function of exponential type π.
2′. $|F_j(x^j)| \geqq (1/e)|x^j|^j$.

3′. $|F_j(x)| \leqq |x^j|^j$ for x real.

4′. $|F_j(x)| \leqq 1$ for $x \in R$, $|x - x^j| \geqq j \log |x^j|$.

Call $B = \{F_j\}$. Then the set B is not bounded in $\mathcal{E}'$ because of condition 2′. Consider $JB = \{JF_j\}$; I claim this set is bounded in $\mathcal{E}'$. Condition 1′ shows that all JF_j are of bounded exponential type. The fact that JB is bounded will be a consequence of the fact that

$$(2.22) \qquad |J(x)F_j(x)| \leqq 1 + |x| + |J(x)| \text{ for } x \in R,$$

as follows from our characterization of the bounded sets of $\boldsymbol{E}'$ (see [11]). To prove (2.22) we consider first those x for which $|x - x^j| \geqq j \log |x^j|$. For such x, inequality (2.22) is an immediate consequence of 4′. On the other hand, if $|x - x^j| \leqq j \log |x^j|$, then 3′ shows that $|F_j(x)| \leqq |x|^{-j}$. By our assumption α, $|x^j| \geqq e^{3j}$ so that if $|x - x^j| \leqq j \log |x^j|$, then $|x| \geqq \frac{1}{2}|x^j|$. Thus, for such x, we have $J(x) \leqq 2^j |x^j|^j$ so that

$$(2.23) \qquad |J(x)F(x)| \leqq 2^j \text{ for } |x - x^j| \leqq j \log |x^j|.$$

But, for such x, $2^j < |x|$ so (2.23) implies (2.22) which completes the proof that (e) *implies* (a) in case $n = 1$.

In case $n > 1$, we proceed exactly as above except that we replace the functions H_j used above by $H'_j(z) = H_j(z_1)H_j(z_2) \cdots H_j(z_n)$. We then define functions F'_j in terms of H_j exactly the way F_j was defined in terms of H_j above. The proof is then concluded exactly as in the case $n = 1$. This completes the proof of Theorem 2.2 and hence of Theorem I.

By a slight modification of the argument used in the proof that (e) *implies* (a), we can prove

PROPOSITION 2.3. *The conditions of Theorem* 2.2 *are equivalent to*

(f) *Let* $U \subset \mathfrak{D}$ *and* $S * U$ *be bounded in* $\mathfrak{D}$; *then* U *is bounded in* $\mathcal{E}'$.

PROPOSITION 2.4. *The conditions of Theorem* 2.2 *imply*

(g) $S * \mathcal{E}'$ *is bornologic.*

(h) $S * \mathfrak{D}$ *is bornologic.*

Proof. We prove the part of the theorem for $\mathfrak{D}$, as the part of the theorem concerning $\mathcal{E}'$ is handled similarly, except that we have to use the fact that condition (d) of Theorem 2.2 implies that $S * U \to U$ is a continuous map of $S * \mathcal{E}' \to \mathcal{E}'$ (see Proposition 2.7 below).

Assume that S satisfies the conditions of Theorem 2.2; let L be a linear function on $S * \mathfrak{D}$ which is bounded on the bounded sets. Define T on $\mathfrak{D}$ by

$$T \cdot h = L \cdot S * h \text{ for } h \in \mathfrak{D}.$$

Then by Theorem 2.2, T is bounded on the bounded sets of $\mathfrak{D}$; since $\mathfrak{D}$ is bornologic, T is continuous on $\mathfrak{D}$, that is, T is a distribution.

Now, let $S * h \to 0$ in the topology of $\mathfrak{D}$; by Theorem 2.2 it follows that $h \to 0$ in the topology of $\mathfrak{D}$. Thus, $T \cdot h \to 0$ so L is continuous on $S * \mathfrak{D}$. This proves that $S * \mathfrak{D}$ is bornologic.

Remark. I do not know if (g) and (h) are true for any S (not necessarily invertible) or whether they imply S is invertible.

THEOREM 2.5. *The conditions of Theorem* 2.2 *are equivalent to*

(i) $S * \mathfrak{D}' \supset \mathfrak{D}$.

Proof. It is clear that (c) implies (i). Assume then that S is not invertible; I shall construct an $f \in \mathfrak{D}$ which is not in $S * \mathfrak{D}'$.

Let us note the following: Let B be a set in $\mathfrak{D}$ for which $S * B$ is bounded in $\mathfrak{D}$. Then if $S * W = f$ for $W \in \mathfrak{D}'$, it must be the case that

$$W \cdot S * h = S * W \cdot h = f \cdot h$$

is uniformly bounded for $h \in B$. Thus, to prove Theorem 2.5 we must produce a set $B \subset D$ with $S * B$ bounded in $\mathfrak{D}$ but $\{f \cdot h\}_{h \in B}$ not bounded. Proposition 2.3 shows that there is hope for this because we can choose B not bounded in $\mathcal{E}'$ with $S * B$ bounded in $\mathfrak{D}$. However, B being not bounded in $\mathcal{E}'$, it follows (see [4]) that B is not weakly bounded in $\mathcal{E}'$. Hence, there exists an $f' \in \mathcal{E}$ so that $\{f' \cdot h\}_{h \in B}$ is not bounded. But the functions $h \in B$ have their carriers in a fixed compact set $K \subset \mathfrak{D}$. Hence, if $f'' \in \mathfrak{D}$ is 1 on K, for any $h \in B$ we have $f' \cdot h = f' f'' \cdot h$. But $f = f' f'' \in \mathfrak{D}$; hence we have $\{f \cdot h\} = \{f' \cdot h\}$ is unbounded which concludes the proof of the theorem.

Remark. Using the notations of the last part of the proof of Theorem 2.2 above ((e) *implies* (a)), we could also write f explicitly (or rather its Fourier transform F) in the form

$$F(z) = \sum c_j H(z - x^j) F'_j(z),$$

where $H \in D$, $H(0) = 1$, $0 \leqq H(x) \leqq 1$ for $x \in R$, and where the c_j are suitably chosen constants.

Theorem 2.5 settles a problem of the author (see [7]), namely, that for $f \in \mathfrak{D}$, $f * \mathfrak{D}' \neq E$. However, I do not know if $\mathfrak{D} * \mathcal{E} = \mathcal{E}$ or even if $\mathfrak{D} * \mathfrak{D}' = \mathcal{E}$, although even $\mathfrak{D} * \mathcal{E} = \mathcal{E}$ is undoubtedly true.

Another interesting question in this connection was raised by Professor

Chevalley in his lectures on the theory of distributions: Is $\mathfrak{D} * \mathfrak{D} = \mathfrak{D}$? This problem seems very difficult. (See appendix at end of paper.)

THEOREM 2.6. *The conditions of Theorem 2.2 are equivalent to*

(j) *For any entire function G, if $JG \in \mathbf{D}$ (or $JG \in \mathbf{E}'$), then $G \in \mathbf{D}$ (resp. $G \in \mathbf{E}'$).*

In fact, for (j) *to hold it is sufficient that $JG \in \mathbf{D}$ should imply $G \in \mathbf{E}'$.*

Proof. If J is slowly decreasing, then by applying the minimum modulus theorem in the manner used in proving Theorem 2.2, (a) *implies* (b), we can show that (j) holds.

Conversely, suppose that S is not invertible; I shall produce an entire function G (which is necessarily of exponential type) such that $JG \in \mathfrak{D}$ but $G \notin \mathbf{E}'$. For this purpose we revert to the notation of the proof of Theorem 2.2, (e) implies (a). Let $H \in \mathbf{D}$ be so chosen that $H(0) = 1$, $0 \leqq H(x) \leqq 1$ for $x \in R$. I want to write first

$$G(z) = \sum c_j H(z - x^j) F'_j(z) \tag{2.24}$$

for suitable constants c_j. Now, it is clear from the construction that I can assume that the x^j are chosen so large that the intervals

$$\{x \mid x \in R, \mid x - x^j \mid \leqq j \log \mid x^j \mid\} \tag{2.25}$$

do not overlap. Then we choose $c^j = j^{-2} \mid x_j \mid^{-j/2}$.

It is clear from (2.24) and (2.25) that the series for G' converges uniformly on the compact sets of R; moreover, for any j,

$$\begin{aligned} G(x^j) &= \sum c_k F'_k(x^j) \\ &\geqq F'_j(x^j) - \sum_{k \neq j} c_k F'_k(x^j) \\ &\geqq (1/ej^2) \mid x^j \mid^{j-j/2} - \sum p^{-2} \\ &\geqq (1/2e) \mid x^j \mid^{j/2} \end{aligned} \tag{2.26}$$

for j sufficiently large because of condition α on the choice of the x^j. Thus, if G is an entire function, it is certainly not in $\mathbf{E}'$.

Next, call $K(z) = (\sin z_1/z_1)(\sin z_2/z_2) \cdots (\sin z_n/z_n)$. For l sufficiently large, $J'(z) = K^l(z) J(z)$ is bounded on R. Since K^l is clearly slowly decreasing, it is sufficient by means of the proof of the first part of this theorem (i.e., that S invertible implies (j)) to prove our result for J' in place of J; that is, we may assume J is bounded on R.

By use of the method of proof of inequality (2.22) above, it follows that

there exists an $a > 0$ so that for any j,

$$\sup_{x \in R} c_j \,|\, F'_j(x) J(x) \,| \leqq a/2^j. \tag{2.27}$$

Since all F'_j are entire functions of exponential type $\leqq \pi$, this shows (see [11]) that the series

$$\sum c_j J(F'_j(z) H(z - x^j) \tag{2.28}$$

converges in the topology of $\boldsymbol{E}'$. It is easy to see, in fact, using the characterization of the topology of $\boldsymbol{D}_l$ that this series converges in $\boldsymbol{D}_l$ for l large enough if the x^j are sufficiently large.

Thus, in particular, the series (2.28) converges in the topology of the space H_l of entire functions on C of exponential type $\leqq i$ for l large enough (see [8]). Using the minimum modulus theorem it follows easily that the series $\sum c_j F_j(z) H(z - x^j)$ also converges in H_l for l large enough. Thus, G is an entire function of exponential type.

In resumé, G is an entire function which is not in $\boldsymbol{E}'$, but $JG \in \boldsymbol{D}$. This completes the proof of Theorem 2.6.

PROPOSITION 2.7. *The conditions of Theorem 2.2 are equivalent to*

(k) $S * W \to W$ *is a continuous linear map of* $S * \mathcal{E}' \to \mathcal{E}'$.

(l) $S * \mathcal{E} = \mathcal{E}$.

Proof. The equivalence of (d), (k), and (l) is a fairly simple consequence of the Hahn-Banach and closed graph theorems and was established by Malgrange in [21].

In all the above we were concerned with invertible operators for the spaces $\mathcal{E}$ and $\mathfrak{D}'$; we wish here to give the analogous description for the space $\mathfrak{D}'_F$ (see [5], [21]). Let $J \in \boldsymbol{E}'$; J is called *very slowly decreasing* if there exists an $A > 0$ so that for any $x \in R$ we can find a $y \in R$ with $|y - x| \leqq A$ and $|J(y)| \geqq (A + |x|)^{-A}$. Then we have

THEOREM 2.2*. *The following conditions are equivalent for* $S \in \mathcal{E}'$:

(a*) *J is very slowing decreasing.*

(b*) $S * f \to f$ *is a continuous linear map of* $S * \mathfrak{D}_F \to \mathfrak{D}_F$.

(c*) $S * \mathfrak{D}'_F = \mathfrak{D}'_F$.

(d*) *There exists an* $S' \in \mathcal{E}'$ *which is invertible for* $\mathfrak{D}'_F$ *such that* $S * T = S'$ *has a solution* $T \in \mathfrak{D}'_F$.

(e*) *For each* $m > 0$ *there is an* $r > 0$ *so that if* B *is a subset of* $\mathfrak{D}$ *for which* $S * B$ *is bounded in* $\mathfrak{D}^r$, *then* B *is bounded in* $\mathfrak{D}^m$.

The proof of Theorem 2.2* is very similar to the proof of Theorem 2.2 and so will be omitted. The equivalence of (c*) and (d*) was proven by Malgrange in [21].

Remark. I have not been able to construct an $S \in \mathcal{E}'$ which is invertible for $\mathcal{D}'$ but not for $\mathcal{D}'_F$.

All the above has concerned itself with the solution of the question of when $S * \mathcal{D}' = \mathcal{D}'$ (or $S * E = E$, etc.). We could ask the question as to what is $S * \mathcal{D}'$ even in case S is not invertible. As mentioned in the introduction, this involves describing explicitly the topology τ on $\mathcal{D}$ which is defined by: N is a neighborhood of zero in τ if $S * N$ is a neighborhood of zero in $S * \mathcal{D}$. That is, τ is the strongest topology so that the map $S * f \to f$ of $\mathcal{D} \to \tau$ is continuous. Of course, in case S is invertible, then τ coincides with $\mathcal{D}$.

Actually, we are not able to give a "good" description of τ; this seems to be because we have not been able to prove that the spaces $S * \mathcal{D}$ are bornologic even if S is not invertible. However, we shall give instead the description of the restriction of τ to each $S * \mathcal{D}_l$. We denote this restriction again by τ.

Of course, we want another expression for the topology τ, one which does not depend so much on S, and one which is useful. We shall give instead the topology σ of the Fourier transform of τ in a form which will be suitable for our purposes. For this we define functions $M_l(z)$ on C which are certain majorants of $J(z)$. For any $z \in C$ and any $l > 5$ (*exponential type* J) we set

$$M_l(z) = \max_{z' \in C; |\mathfrak{J}(z')| \leq \mathfrak{J}(z)} \exp(-l\,|z' - z|)\,|J(z')|. \tag{2.29}$$

We shall describe the topology σ using M_l instead of J. This is a great advantage because M_l behaves much more regularly than J, and the zeros of J do not enter into the M_l.

THEOREM 2.8. *For each $l' > 0$ the topology σ on $\mathcal{D}_{l'}$ can be described as follows: Let l be a fixed number $> l' + d$, where d is some number depending only on the exponential type of J. For each integer $m > 0$ call N the set of $F \in \mathcal{D}_l$ such that*

$$|z^m{}_k F(z)| \leq \exp(l\,|\mathfrak{J}(z)|) \tag{2.30}$$

for $k = 1, 2, \cdots, n$. Then these sets form a fundamental system of neighborhoods of zero for σ.

Proof. As in the proof of Theorem 2.6 we may assume $|J(x)| \leq 1$ for

$x \in R$. Since $M_l(z) \geqq J(z)$ for all z, we have only to prove that the sets N are neighborhoods of zero for τ.

Let $z \in C$ be fixed; let z' be a point with $\mathfrak{I}z' = \mathfrak{I}z$, where $M_l(z) = \exp(-l\,|z'-z|)\,|J(z')|$. We shall assume first that $n=1$. Suppose that we had

$$(2.31) \qquad |z''^m F(z'') J(z'')| \leqq \exp(l'\,|I(z'')|$$

for all $z'' \in C$. Now, by the minimum modulus theorem, we can draw about the point z' a circle γ of radius between $|z'-z|$ and $2\,|z'-z|$ so that for all $z'' \in \gamma$ we have for certain constants c and d which depend only on the exponential type of J,

$$(2.32) \qquad |J(z'')| \geqq c\exp(-d\,|z'-z|)\,|J(z')|\exp(-d\,|\mathfrak{I}z|).$$

Combining (2.31) and (2.32) we have for all $z'' \in \gamma$,

$$(2.33) \qquad |J(z')F(z'')z''^m| \leqq (1/c)\exp(d\,|z'-z| + (l'+d)\,|\mathfrak{I}z|).$$

Since $F(z'')z''^m$ is an entire function of z'' we have, by the maximum modulus theorem,

$$(2.34) \qquad |J(z')F(z)z^m| \leqq (1/c)\exp(d\,|z'-z| + (l'+d)\,|\mathfrak{I}z|)$$

or, what is the same thing,

$$(2.35) \qquad |J(z')\exp(-d\,|z'-z|)F(z)z^m| \leqq (1/c)\exp[(l'+d)\,|\mathfrak{I}z|].$$

Now, if l is larger than d,

$$(2.36) \qquad M_l(z) = \exp(-l\,|z'-z|)\,|J(z')| \leqq \exp(-d\,|z'-z|)\,|J(z')|.$$

Thus (2.35) implies

$$(2.37) \qquad |M_l(z)F(z)z^m| \leqq (1/c)\exp[(l'+d)\,|\mathfrak{I}z|]$$

which gives our result in case $n=1$.

The case $n > 1$ is handled by the same method except that we apply the minimum and maximum modulus theorem in each variable separately. We shall omit the details.

All that has been done previously in regard to the invertible operators is in connection with the question of when $S * \mathfrak{D}' = \mathfrak{D}'$. On the other hand, we might ask when does $S * \mathfrak{D}' \supset T * \mathfrak{D}'$, where T is a distribution in $\mathcal{E}'$? Call J the Fourier transform of S and K the Fourier transform of T. We might expect that $S * \mathfrak{D}' \supset T * \mathfrak{D}'$ should be equivalent to the fact that K/J does not tend to zero too fast at infinity. However, we are not able to establish

this fact; this problem seems to be essentially the same as the problem of describing the topology σ on $\boldsymbol{D}$ itself which, as we mentioned, we are not able to accomplish. However, we can prove part of the analogue for the spaces D'_l. For this purpose we make the following

Definition. We say that J/K is *slowly decreasing* if for each l sufficiently large there exists a j so that for all $z \in C$,

$$M_l(K;z)/M_l(J;z) \leqq j(1+|\mathcal{R}z|)^j \exp(d\,|\,\mathcal{I}z\,|), \tag{2.38}$$

where $d = 100n$ (*exp. type* J + *exp. type* $K + 1$). (Here we have written $M_l(J;z)$, $M_l(K;z)$ to avoid confusion.)

It is easily seen that $J/1$ is slowly decreasing in the above sense if and only if J is slowly decreasing in our previous sense.

We can now formulate a partial extension of Theorem 2.2:

THEOREM 2.9. *For $S, T \in \mathcal{E}'$ each property implies the succeeding one:*

(a′) *J/K is slowly decreasing.*

(b′) *$S * f \to T * f$ is a continuous linear map of $S * \mathcal{D}_k \to T * D_k$ for each k.*

(c′) *For each k there exists a k' so that $S * \mathcal{D}'_{k'} \supset T * \mathcal{D}'_k$.*

(d′) *For each k there exists a $W \in \mathcal{D}'_k$ such that $S * W = T$.*

(e′) *$S * U \to T * U$ is a semi-continuous linear map of $S * \mathcal{E}' \to T * \mathcal{E}'$.*

Proof. As in the proof of Theorem 2.2, the implications (b′) *implies* (c′), (c′) *implies* (d′), and (d′) *implies* (e′) are easy. Moreover, (a′) *implies* (b′) is an easy consequence of Theorem 2.8.

We are not able to prove (e′) implies (a′). However, we can prove

THEOREM 2.10. *If condition* (e′) *holds, then for any $\epsilon > 0$ we can find a j (possibly depending on ϵ) so that*

$$[M_l(K,z)]^{1+\epsilon}/M_l(J,z) \leqq j(1+|\mathcal{R}z|)^j \exp(d'\,|\,\mathcal{I}z\,|). \tag{2.39}$$

(Here $d' = d + \pi/\epsilon$.)

Proof. Assume (e′) holds but (2.39) does not hold. Then we can find an $\epsilon > 0$ and a sequence of points z^j with $z^j \to \infty$ fast enough so that

$$[M_l(K,z^j)]^{1+\epsilon}/M_l(J,z^j) > j(1+|\mathcal{R}z^j|)^j \exp(d'\,|\,\mathcal{I}z\,|). \tag{2.40}$$

As in the proof of Theorem 2.6 we may assume $|K(x)| \leqq 1$ and $|J(x)| \leqq 1$ for $x \in R$.

We shall show first that we may assume that z^j are so chosen that $M_l(K, z^j) = |K(z^j)|$. For this purpose, assume (2.40) holds and let w be chosen so that $M_l(K, z^j) = \exp(-l\,|z^j - w|)\,|K(w)|$. Then clearly

$$(2.41) \qquad M_l(K, w) \geqq K(w) = \exp(l\,|z^j - w|)\,M_l(K, z^j).$$

I claim that we must have $M_l(J, w) \leqq \exp(l\,|z^j - w|)\,M_l(J, z^j)$. For, assume this is not the case; let v be chosen so that $M_l(J, z^j) = \exp(-l\,|z^j - v|)\,J(v)$. Then if $M_l(J, w) > \exp(l\,|z^j - w|)\,M_l(J, z^j)$, we would have by the triangle inequality

$$\begin{aligned} M_l(J, w) &> \exp(l\,|z^j - w|)\exp[-l(|z^j - v|)]\,J(v) \\ &\geqq \exp(-l\,|w - v|)\,J(v) \end{aligned}$$

which contradicts the definition of $M_l(J, w)$. Moreover, this argument shows that we have equality in (2.41).

Thus we have shown that

$$[M_l(K, z^j)]^{1+\epsilon}/M_l(J, z^j) \leqq ([M_l(K, w)]^{1+\epsilon}/M_l(J, w))\exp(-\epsilon l\,|z^j - w|)$$

which implies that (2.40) holds with w in place of z^j. Hence, we may assume that z^j satisfies

$$(2.42) \qquad M_l(K, z^j) = |K(z^j)|.$$

Next, I want to obtain an estimate for a cube containing z^j so that for all points z with $\mathfrak{J} z = \mathfrak{J} z^j$ in this cube we have

$$|J(z)| \leqq |K(z^j)|\,j^{-1/2}\exp(-d'\,|\mathfrak{J} z^j|/2)\,(1 + |\mathfrak{R} z^j|)^{-j/2}.$$

Let w be a point for which $\mathfrak{J} w = \mathfrak{J} z^j$ and

$$|J(w)| > |K(z^j)|\,j^{-1/2}\exp(-d'\,|/2)\cdot(1 + |\mathfrak{R} z^j|)^{-j/2}.$$

Then we estimate $M(J; z^j)$ as follows:

$$\begin{aligned} M_l(J; z^j) &\leqq \exp(-l\,|z^j - w|)\,|J(w)| \\ &> \exp(-l\,|z^j - w|)K(z^j\,|\,j^{-1/2}\exp(-d'\,|\mathfrak{J} z^j|/2)\,(1 + \mathfrak{R} z^j|)^{-j/2}. \end{aligned}$$

On the other hand, we know by (2.40) that

$$M_l(J; z^j) < |K(z^j)|^{1+\epsilon}/j(1 + |\mathfrak{R} z^j|)^j\exp(d'\,|\mathfrak{J} z^j|).$$

Thus we must have

$$|K(z^j)|\exp(-l\,|z^j - w|) < |K(z^j)|^{1+\epsilon}/j^{1/2}(1 + |\mathfrak{R} z^j|)^{j/2}\exp(d'\,|\mathfrak{J} z^j|/2).$$

Hence,

$$-l\,|\,z^j - w\,| < \epsilon \log |\,K(z^j)\,| - \tfrac{1}{2}\log j - (j/2)\log(1+|\,\mathcal{R}z^j\,|) - d'\,|\,\mathcal{I}z^j\,|/2.$$

or

$$|\,z^j - w\,| > -(\epsilon/l)\log|\,K(z^j)\,| + \tfrac{1}{2}\log j + (j/2)\log(1+|\,\mathcal{R}z^j\,|) + d'\,|\,\mathcal{I}z^j\,|/2. \tag{2.43}$$

As in the proof of Theorem 2.2, (e) *implies* (a), the proof for general n is similar to the proof for $n=1$ which we shall henceforth assume. We shall also use the notation of the proof of Theorem 2.2, (e) *implies* (a). We may clearly assume for simplicity that $\mathcal{I}z^j \geqq 0$ for all j and that $0<\epsilon<1$. Then we set

$$F''_j(z) = e^m H_m((z-z^j)/\epsilon)\exp[-i(\pi+1)(z-z^j)/\epsilon], \tag{2.44}$$

where

$$m = [-\log|\,K(z^j)\,| + d'\,|\,\mathcal{I}z^j\,|/2 + \tfrac{1}{2}\log j + (j/2)\log(1+|\,\mathcal{R}z^j\,|)], \tag{2.45}$$

the bracket denoting as usal the integral part. Note that since $|\,K(x)\,| \leqq 1$ for $x\in R$, $-\log|\,K(z^j)\,| + d'\,|\,\mathcal{I}z^j\,|/2 > 0$. Then we have the following properties:

1″. F''_j is an entire function of exponential type $2(\pi+1)/\epsilon$.

2″. $|\,K(z^j)F''_j(z^j)\,| \geqq (1/e)j^{1/2}(1+|\,\mathcal{R}z^j\,|^{j/2})\exp(d'\,|\,\mathcal{I}z^j\,|/2)$.

3″. $|\,F''_j(z)\,| \leqq |\,K(z^j)\,|^{-1}\exp(d'\mathcal{I}z^j/2)\,(j/2)\,(1+|\,\mathcal{R}z^j\,|)^{j/2}$ for $Iz = Iz^j$.

4″. $|\,F''_j(z)\,| \leqq 1$ for $\mathcal{I}z = \mathcal{I}z^j$,

$$|\,z-z^j\,| \geqq \epsilon(-\log|\,K(z^j)\,| + d\,|\,\mathcal{I}z^j\,|/2 + \tfrac{1}{2}\log j + (j/2)\log(1+|\,\mathcal{R}z^j\,|). \tag{2.46}$$

Conditions 1″ and 2″ show that $\{KF''_j\}$ is not bounded in $\mathcal{E}'$. Condition 3″ together with the argument following (2.42) shows that $|\,J(z)F''_j(z)\,| \leqq 1$ for $\mathcal{I}z = \mathcal{I}z^j$ and

$$|\,z-z^j\,| \leqq -(\epsilon/l)\log|\,K(z^j)\,| + \tfrac{1}{2}\log j + (j/2)\log(1+|\,\mathcal{R}z^j\,|) + d'\,|\,\mathcal{I}z^j\,|/2. \tag{2.47}$$

Condition 4″ shows that for z satisfying (2.46) we have

$$|\,J(z)F''_j(z)\,| \leqq \exp(d_0\,|\,\mathcal{I}z^j\,|) \qquad (d_0 = \text{exp type } J).$$

Hence, by the Phragmén-Lindelöf theorem, we have for all $x\in R$ (since we may assume $\epsilon^{-1} > d_0 = \text{exp type } J$) $|\,J(x)F''_j(x)\,| \leqq 1$. This proves that $\{JF''_j\}$ is bounded in $\mathcal{D}_k$ for k large enough, which concludes the proof of Theorem 2.10.

Remark 1. By a slight modification of the above process we could prove that if (2.39) does not hold, then $S*f\to T*f$ is not a semi-continuous map of $\mathfrak{D}$ into $\mathcal{E}'$.

Remark 2. In case $K=1$, of course, we can replace $[M(K;z)]^{1+\epsilon}$ by $[M(K;z)]$ because $M(K;z)$ is itself $>(j+|\mathcal{R}z|^{-j}\exp(-d|\mathcal{I}z|)$ for some j. Thus the conditions (a′), (b′), (c′), (d′), (e′) are equivalent (and this fact is, of course, much weaker than Theorem 2.2). But the arguments used above can show that in this case, if there exists a distribution $W\in\mathfrak{D}'$ with $S*W=\delta$ on $\mathfrak{D}_k$, then S is invertible.

Remark 3. We have taken only one possible choice for the majorants M_l; actually, many possibilities present themselves. For example, we could replace the right side of (2.29) by

$$M_l(J,\lambda;z)=\max_{z'\in C,\,|\mathcal{I}(z')|\leqq|\mathcal{I}(z)|+\lambda}\exp(-l\,|z'-z|)\,|J(z')|, \tag{2.48}$$

where $\lambda\geqq 0$ is suitably chosen. All the above theorems would then be proven with no essential modification. The above generalization has the advantage that $(n=1)$ the ratio of majorants

$$M_l(J',\lambda;z)/M_l(J,\lambda+1;z)\leqq e^l \tag{2.49}$$

for any J. This follows immediately from the above and Cauchy's formula for the derivative of an analytic fuction. The result (2.49) seems of great importance in understanding the deeper parts of the theory of mean periodic functions for it shows that (in a slightly broader sense than used above) J/J' is slowly decreasing. (Compare Section 3 below and [27], particularly the latter where the properties of the ratio J'/J are of great importance.).

3. Invertibility and mean periodic functions. We are now going to study the relationship of the above with L. Schwartz' theory of mean periodic functions (see [25], [26], [27], [8]). We assume that $n=1$ in the following because Schwartz' mean periodic expansion holds only in this case except for some special cases in higher dimension (see [15], [16]). We shall first briefly recall the main aspects of this theory:

Let V be a closed linear subset of $\mathcal{E}$ which is closed under translation (and hence also convolution by elements of $\mathcal{E}'$). We want to expand a given function $f\in V$ in terms of the exponential polynomials which belong to V; in particular, we wish to show that every $f\in V$ is the limit of the exponential polynomials of V. Assume V is not all of $\mathcal{E}$; then there exists an $S\in\mathcal{E}'$ satisfying $S\cdot f=0$ for all $f\in V$. Since V is closed under translation we also

have $S * f = 0$ for all $f \in V$. In particular, the Fourier transform J of S must vanish at each point $z \in C$ for which $\exp(iz\cdot) \in V$ and J must have a zero at z of order at least $j+1$ if $x^j \exp(iz\cdot x) \in V$.

Now, suppose we had some expansion

$$f(x) \sim \sum_{j,k} c_{jk} x^j \exp(iz^k \cdot x), \tag{3.1}$$

where the sum is taken over all pairs j, k for which $x^j \exp(iz^k \cdot x) \in V$.

We denote by $\{1/J(z)\}_{z^k}$ the principal part of the expansion of $1/J(z)$ at z^k; we set $J_k(z) = J(z)\{1/J(z)\}_{z^k}$, and we call S_k the Fourier transform of J_k. Then a simple computation shows that if (3.1) holds and if we have some kind of convergence, then (see [27])

$$(S_k * f)(x) = \sum_{j=0}^{j'-1} c_{jk} x^j \exp(iz^k \cdot x), \tag{3.2}$$

where j' is the order of the zero of S at z^k. Thus, formally,

$$f = \sum f * S_k = f * \sum S_k \tag{3.3}$$

provided the sum $\sum S_k$ converges in an appropriate sense. Now, (3.3) would hold for all $f \in V$, or even for all f which satisfy $S * f = 0$ provided that we could demonstrate the existence of a $T \in \mathcal{E}'$ with

$$\delta = \sum S_k + S * T. \tag{3.4}$$

The Fourier transform of this relationship is

$$1 = \sum J\{1/J\}_{z^k} + JK, \tag{3.5}$$

where K is the Fourier transform of T. Let us note the following:

$$(d^p/dz^p)[J\{1/J\}_{z^k}](z^l) = \begin{cases} 0 \text{ if } k \neq l, \ 0 \leqq p \leqq j'_l + 1 \\ 1 \text{ if } k = l, \ p = 0 \\ 0 \text{ if } k = l, \ 1 \leqq p \leqq j'_k + 1 \end{cases} \tag{3.6}$$

because $1 = J\{1/J\} + J \cdot$ regular part. Thus it follows that J divides $1 - \sum J\{1/J\}_{z^k}$ in the ring of entire functions. Now, suppose we can show $\sum J\{1/J\}_{z^k}$ belongs to $\boldsymbol{E}'$; then if J is slowly decreasing, the existence of K will be verified by Theorem 2.6.

Thus, if J is slowly decreasing, the whole theory of mean periodic functions will have a very simple structure. In case J is not slowly decreasing then it does not seem that formulae like (3.1) or (3.3) can hold; rather we shall show that they hold only in a certain limit sense.

Our main job is to show:

Suppose J is slowly decreasing. Then it is possible to find groupings $G_1, G_2, \cdots$ of the points z^k so that the series

$$\sum_{r=1}^{\infty} \sum_{z^k \in G_r} J\{1/J\}_{z^k} \tag{3.7}$$

converges in the topology of E'.

This statement is not quite true (or, at least, I cannot prove it), and we shall derive a slightly modified form (Theorem 3.1 below). Following the method of Schwartz (see [25]) we write for $z \neq z^k$

$$(\{1/J\}_{z^k})(z) = \int_{\Gamma_k} d\zeta / J(\zeta)(z-\zeta),$$

where Γ_k is a closed curve containing z^k in its interior but not containing z or any $z^{k'}$ for $k' \neq K$. Hence, if $z \neq z^k$ for any $z^k \in G_r$,

$$\sum_{z^k \in G_r} \{1/J\}_{z_k} = \int_{\Gamma_r} d\zeta / J(\zeta)(z-\zeta), \tag{3.8}$$

where now Γ_r is a closed curve containing all $z^k \in G_r$ but not containing z or any $z^{k'} \notin G_r$.

The fact that z does not lie in any Γ_r is of no consequence for the convergence of (3.7) because if z lies in Γ_r, then we would get a contribution of $1/J(z)$ (if $J(z) \neq 0$) to the integral in (3.8). Since we are going to multiply by $J(z)$ anyway, this does not affect the convergence. Thus, I want to find a sequence of groupings G_r so that I can prove the series

$$\sum_{r=1}^{r} \int_{\Gamma_r} d\zeta / J(\zeta)(z-\zeta) \tag{3.9}$$

converges in a suitable sense. It is clear that the contours Γ_r have to be chosen in such a way that J is large on Γ_r; the possibility of choosing such Γ_r depends on the fact that J is slowly decreasing.

I shall assume first that all the zeros of J are real; we shall explain later how the former restriction is removed. Now, since J is slowly decreasing, there exists a positive integer j so large that for each $x \in R$ there is a $y \in R$ with $|y-x| \leqq j \log(j+|x|)$ and $|J(y)| \geqq (j+|y|)^{-j}$. For each integer k, positive or negative or zero, let A_k be the interval

$$A_k = \{x \in R \mid |x-k| \leqq j \log(j+|k|+2)\} \tag{3.10}$$

By the above, in A_k there is a point y_k with $|J(y_k)| \geqq (j+|y_k|)^{-j}$.

Now we are in a position to apply the minimum modulus theorem to

construct Γ_k. About y_k we can draw a circle Γ'_k of center y_k, radius R_k such that

$$(3.11) \qquad 4j\log(j+|k|+2) \leqq R_k \leqq 8j\log(j+|k|+2),$$

so that for all points $z \in \Gamma'_k$ we have

$$(3.12) \qquad |J(z)| \geqq (l+|y_k|)^{-l}$$

for some $l > 0$ which depends only on J (see Theorem 5 of [8], p. 317). We need a slight sharpening of this estimate: Not only does (3.12) hold for all points on Γ'_k but we can find a number $q > 0$ depending only on J so that (3.12) holds for all points z with

$$(3.13) \qquad R_k - q \leqq |z - y_k| \leqq R_k + q.$$

The proof of this can be obtained by a slight modification of the proof of Theorem 5 of [8], p. 317.

We notice that if we replace J by $J'(z) = J(z)z^{l+d}$ (d sufficiently large) then inequality (3.12) can be improved to

$$(3.14) \qquad J'(z) \geqq c(1+|y_k|)^d$$

for all z satisfying (3.13).

Now, we are ready to define the curves Γ_r (which depend slightly on z). we set $\Gamma_0 = \Gamma'_0$ unless $||z| - |R_0|| < q$ in which case we replace Γ'_0 by a circle of radius R'_k between $R_k - q$ and $R_k + q$ so that $||z| - |R'_k|| < q$. Suppose Γ_r have been defined for $r = 0, \pm 1, \pm 2, \cdots, \pm r'$. Then de define $\Gamma_{r'+1}$ as follows: ($\Gamma_{r'-1}$ is defined similarly)

1. If $y_{r'+1}$ is contained in or on $\bigcup_{|r| \leqq r'} \Gamma_r$, then $\Gamma_{r'+1}$ is empty.

2. If $y_{r'+1}$ is not contained in or on $\bigcup_{|r| \leqq r'} \Gamma_r$ and if $||z - y_{r'+1}| - R_{r'+1}| \geqq q$, then $\Gamma'_{r'+1}$ intersects $\bigcup_{|r| \leqq r'} \Gamma_r$ and we choose that connected component of $\Gamma'_{r'+1}$ *minus* this intersection which meets the real axis at a point $> y_{r'+1}$. $\Gamma_{r'+1}$ is the union of this component with arcs of $\bigcup_{|r'| \leqq r} \Gamma_r$, these arcs being chosen in such a manner that $\Gamma_{r'+1}$ is closed, simple, and does not contain in its interior any points which lie in the interior of some Γ_r for $|r| \leqq r'+1$. It is easily seen that this curve is uniquely determined by our description.

3. If $y_{r'+1}$ is not contained in or on $\bigcup_{|r| \leqq r'} \Gamma_r$ and if $||z - y_{r'+1}| - R_{r'+1}| < q$, then we choose an $R'_{r'+1}$ so that $||z - y_{r'+1}| - R'_{r'+1}| \geqq q$ and $|R'_{r'+1} - R_{r'+1}| \leqq q$ and we proceed as in 2 above.

Now, the number of circles Γ'_r that $\Gamma'_{r'}$ can meet for $|r| < |r'|$ is certainly $< 2r'$. Since R_r satisfies (3.13), it follows easily that the length of $\Gamma_{r'}$, which cannot exceed the sum of the circumferences of circles of radii $R_r + q$, must be $\leqq \text{const}(1 + |r'|)^2$.

We can now prove

LEMMA 3.1. *The series* $\sum_{r=-\infty}^{\infty} \int_{\Gamma_r} d\zeta/J'(\zeta)(z-\zeta)$ *converges uniformly for* $z \in C$.

Proof. Our estimates show that on Γ_r we have $|J'(\zeta)| \geqq c(1 + |y_r|)^d$ because of (3.14) and the fact that $(1 + |y|)^d$ is monotonic in $|y|$. The length of Γ_r is $\leqq c'(1 + |r|)^2$. Moreover, by construction, for $\zeta \in \Gamma_r$ we have $|z - \zeta| \geqq q$. Lemma 3.1 follows immediately if d is sufficiently large because the number of y_r with $|y_r| \leqq |k|$ is by construction less than

$$|k| + j\log(j + |k| + 2).$$

Now, we note that by Cauchy's theorem and the definitions, the integrals $\int_{\Gamma_r} d\zeta/J'(\zeta)(z-\zeta)$ do not depend on z except for the term $1/J(z)$ which depends on whether z lies inside or outside Γ_r. Thus,

$$(3.15) \quad \sum J'(z) \int_{\Gamma_r} d\zeta/J'(\zeta)(z-\zeta) = \begin{cases} 1 + \sum\limits_{r} \sum\limits_{z^k \in G_r} J'_k(z) & \text{if } \epsilon \text{ lies in some } \Gamma_r \\ \sum\limits_{r} \sum\limits_{z^k \in G_r} J'_k(z) & \text{otherwise.} \end{cases}$$

The series on the left of (3.15) obviously converges in the topology of $\boldsymbol{E}'$ (hence, so does the right side), where we have written G_r for all those z^k contained in Γ_r.

Thus we have shown that $\sum\limits_{r} \sum\limits_{z^k \in G_r} J'_k(z)$ converges in $\boldsymbol{E}'$; as we have noted above (see (3.6) and following) this means we can write

$$(3.16) \quad \begin{aligned} 1 &= J'(z)K'(z) + \sum_{r} \sum_{z^k \in G_r} J'_k(z) \\ &= J(z)K(z) + \sum_{r} \sum_{z^k \in G_r} J'_k(z), \end{aligned}$$

where $K(z) = z^d K'(z)$.

Now, we shall show how to eliminate the restriction that J have only real zeros. By slightly modifying our above constructions we can show that we can construct three sequences of contours:

1. $\{\gamma_k\}$ in a manner similar to $\{\Gamma_k\}$ above
2. $\{\gamma'_k\}$ in the upper half plane
3. $\{\gamma''_k\}$ in the lower half plane

in such a manner that each zero of J is contained in exactly one γ_k, and, if we call $\mathfrak{J}'(z)=z^dJ(z)$, $\mathfrak{J}''(z)=\exp(-idz)J'(z)$, $\mathfrak{J}'''(z)=\exp(idz)J'(z)$, then, for d large enough, the three series

$$\Sigma\int_{\gamma_k} d\zeta/\mathfrak{J}'(\zeta)(z-\zeta),\ \ \Sigma\int_{\gamma'_k} d\zeta/\mathfrak{J}''(\zeta)(z-\zeta),\ \ \Sigma\int_{\gamma''_k} d\zeta/\mathfrak{J}'''(\zeta)(z-\zeta)$$

converge uniformly for $z\in C$. If we denote by G'_r the set of $z^k\in\gamma_r$, G''_r the set of $z^k\in\gamma'_r$, and z''_r the set of $z^k\in\gamma''_r$, then the above shows that the three series

$$\sum_r\sum_{z^k\in\gamma'_r}\mathfrak{J}'(z)\{1/\mathfrak{J}'(z)\}_{z^k},\quad \sum_r\sum_{z^k\in G''_r}\mathfrak{J}''(z)\{1/\mathfrak{J}''(z)\}_{z^k},$$

$$\sum_r\sum_{z^k\in G'''_r}\mathfrak{J}'''(z)\{1/\mathfrak{J}'''(z)\}_{z^k} \tag{3.17}$$

converge in the topology of $\boldsymbol{E}'$. It follows immediately from the definitions that

$$1-\sum_r\sum_{z^k\in G'_r}\mathfrak{J}'(z)\{1/\mathfrak{J}'(z)\}_{z^k}+\sum_r\sum_{z^k\in G''_r}\mathfrak{J}''(z)\{1/\mathfrak{J}''(z)\}_{z^k}$$
$$+\sum_r\sum_{z^k\in G'''_r}\mathfrak{J}'''(z)\{1/\mathfrak{J}'''(z)\}_{z^k}$$

is a function in $\boldsymbol{E}'$ which vanishes at each z^k to the order j'_k; hence, is of the form $K(z)J(z)$ for some $K\in E'$ (by Theorem 2.6). For each k, moreover, we see that $\mathfrak{J}'_k$, or $\mathfrak{J}''_k$, or $\mathfrak{J}'''_k$ is a multiple of J_k. If we now denote by $\{G_r\}$ some ordering of the three sequences $\{G'_r\}$, $\{G''_r\}$, and $\{G'''_r\}$, then we have:

THEOREM 3.1. *Suppose that $S\in\mathcal{E}'$ is invertible. Then we can find a sequence of distributions $T_k\in\mathcal{E}'$ each of which is of the form U_k*S_k with $U_k\in\mathcal{E}'$ so that for some grouping of terms $\{G_r\}$ the series $\sum_r\sum_{z^k\in G_r}T_k$ converges in the topology of $\mathcal{E}'$. We can find a $W\in\mathcal{E}'$ so that*

$$\delta=S*W+\sum_r\sum_{z^k\in G_r}T_k. \tag{3.18}$$

If $f\in\mathcal{E}$ satisfies $S*f=0$, then T_k*f are exponential polynomials which depend only on f (not on S or T_k). Hence, the series

$$\sum_r\sum_{z^k\in G_r}T_k*f \tag{3.19}$$

which converges in $\mathcal{E}$ represents the mean periodic expansion of f in terms of exponential polynomials.

Proof. All has been proven except the statement that S'_k*f are exponential polynomials which depend only on f in case $S*f=0$. First we notice that is is clear that if j''_k denotes the order of the zero of J at z^k

(or d if $z^k = 0$) then $(z - z^k)^{j''_k} L_k(z)$ is a multiple of J in the ring $\mathcal{E}'$, where L_k is the Fourier transform T_k. Thus, $(d/dx - z^k)^{j''_k} T_k * f = 0$, which means that $T_k * f$ is an exponential polynomial.

We note that $T_k * T_l$ is a multiple of $S_k * S_l$ which is a multiple of S for $k \neq l$. Thus, since the series (3.19) converges to f in the topology of $\mathcal{E}$, we have

$$T_l * f = T_l * T_l f,$$

that is, convolution by T_l is an idempotent for the solutions of $S * f = 0$ which is a projection on the exponential polynomial corresponding to z^l.

Thus, if f satisfied an equation $S^1 * f = 0$ and we have an expansion corresponding to (3.18) for S^1:

$$\delta = S^1 * W^1 + \sum_r \sum_{k \in G^1_r} T^1{}_k,$$

then we would have by the above

$$f = \sum_r \sum_{k \in G_r} T_k * T^1{}_k * f, \tag{3.20}$$

where only those k appear which are common zeros of J and J^1. By (3.20) we have

$$T_k * f = T^1{}_k * (T_k * f) \tag{3.21}$$

so $T_k * f$ is an exponential polynomial for which the degree of the polynomial is $\leqq$ the order of the zero of J^1 at j^k. It is easily seen by Fourier transform that $T^1{}_k$ acts as the identity on such exponential polynomials; we have $T_k * f = T^1{}_k * f$, which shows that the $T_k * f$ depend only on f. (The above assumes that the order of the zero of J at z^k is $\geqq$ order of zero of J^1 at z^k; if this is not the case, the roles of J and J^1 are interchanged.)

This completes the proof of Theorem 3.1.

In case J is not slowly decreasing, then the method of Schwartz (see [27]) shows the existence of a sequence of groupings $\{G_r\}$ such that for each $\epsilon > 0$ the series $\sum_r \sum_{k \in G_r} \exp(-\epsilon \mid z^k \mid) S_k$ converges in the topology of $\mathcal{E}'$. Moreoever, $\lim_{\epsilon \to 0} \sum_r \sum_{k \in G_r} \exp(-\epsilon \mid z^k \mid) S_k$ exists in the topology of $\mathcal{E}'$ and its difference from δ is equal to an element of the *closed* ideal generated by S. This gives Schwartz' main result on the convergence of the mean periodic expansion by means of grouping of terms and Abel convergence factors.

Remark 1. I do not know whether the following weak converse of Theorem 3.1 holds; that is: If there exists an identity like (3.8) with T_k of the form $U_k * S_k$ and if the series on the right side converges, then for

some entire function P of exponential type (but possibly not in $\mathcal{E}'$), $PJ \in \mathcal{E}'$ is slowly decreasing.

Remark 2. Theorem 3.1 can be extended so as to apply to distribution solutions V of $S * V = 0$. In fact, the proof is exactly the same as the proof for $f \in \mathcal{E}$.

4. The intersection of $S * \mathfrak{D}'$ for $S \in \mathcal{E}'$. In this section we shall prove Theorem II of the Introduction, that is, that $\bigcap_{S * \mathcal{E}'} S * \mathfrak{D}' = A$ the space of real analytic functions on R. As we mentioned in the Introduction, the proof that $\bigcap_{S * \mathcal{E}'} S * \mathfrak{D}' \subset A$ depends on the Denjoy-Carleman theorem for quasi-analytic functions which we now recall.

Let $\{M_j\} = M$ be a sequence of positive numbers. We define the class A_M as consisting of all functions f which are defined and indefinitely differentiable on the interval $-1 < x < 1$ and satisfy for some $B, K > 0$

$$|f^{(j)}(x)| \leqq BK^j M^j_j \tag{4.1}$$

for all x in this interval. The class A_M is called non quasi-analytic if there exists an $f \in A_M$, $f \neq 0$, such that f and all its derivatives vanish at some point x in the interval $-1 \leqq x \leqq 1$.

Theorem of Denjoy-Carleman. A_M is quasi-analytic if and only if the series

$$\sum_{j=0}^{\infty} (1/\bar{M}_j) \tag{4.2}$$

diverges, where $\bar{M} = \{\bar{M}_j\}$ is the monotonic increasing minorant of M.

From this theorem we deduce the following proposition which will be our key tool in the proof that $\bigcap_{S * \mathcal{E}'} S * \mathfrak{D}' \subset A$:

PROPOSITION 4.1. *Let M be monotonic increasing, with $\sum(1/M_j) < \infty$. Then there exists an $f \in A_M$ which vanishes outside of a compact subset of $-1 \leqq x \leqq 1$ but does not vanish identically.*

Proof. By the Denjoy-Carleman theorem we can find a function $g \in A_M$, $g \neq 0$, which vanishes with all its derivatives at a, say where $-1 < a < 1$. Then g does not vanish identically in at least one of the intervals $-1 < x < a$, $a < x < 1$, we suppose it is the latter. Let b be the greatest lower bound of all $x > a$ for which $g(x) \neq 0$, and set

$$g_1(x) = \begin{cases} g(x) & \text{for } x \geqq b \\ 0 & \text{for } x < b. \end{cases} \tag{4.2}$$

It is clear that g_1 is again in A_M. Finally, it is clear that for ϵ sufficiently small,

$$f(x) = \begin{cases} g_1(x) g_1(2b - 2x + \epsilon) & \text{for } b \leqq x \leqq b + \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

vanishes outside of a compact subset of $-1 < x < 1$ and is not identically zero. Since for $b \leqq x \leqq b + 1$ we have

$$| g_1^{(j)}(2b - x + \epsilon) | \leqq BK^j M^j{}_j,$$

all that remains to prove is that whenever two functions satisfy (4.1) so does their product (with possibly different B, K) if M is monotonic.

Let p, q satisfy (4.1). Then for any j,

$$\begin{aligned} | (pq)^{(j)}(x) | &= \Big| \sum_{k=0}^{j} p^{(k)}(x) q^{(j-k)}(x) (C_{k+1}{}^{j+1}) \Big| \\ &\leqq \sum_{k=0}^{j} B^2 M^k{}_k K^k M^{j-k}{}_{j-k} K^{j-k} (C_{k+1}{}^{j+1}) \\ &\leqq B^2 M^j{}_j K^j \sum_{k=0}^{j} (C_{k+1}{}^{j+1}) \\ &= B^2 M^j{}_j (2K)^j \end{aligned}$$

because M is monotonic which shows that pq satisfies (4.1) for larger K, B. This completes the proof of Proposition 4.1.

THEOREM 4.2. $\bigcap_{S * \mathcal{E}'} S * \mathcal{D}' \subset A$.

Proof. Let $f \in \bigcap_{S * \mathcal{E}'} S * \mathcal{D}'$ and suppose that f is not real analytic; it is clear anyway that $f \in \mathcal{E}$. Suppose for simplicity that f is not analytic in the neighborhood of $x = 0$. First we note that we can prove easily using the theory of elliptic differential operators and this means we can find a sequence of points $\{a_j\}$ with $| a_j | \leqq 1$ and a corresponding sequence of positive integers m_j which are strictly increasing to infinity so that

$$| f_{m_j}(a_j) | \geqq j^{6nm_j} (2m_j!)^n, \tag{4.4}$$

where we have written f_{m_j} for $\Delta^{m_j} f$, Δ denoting the Laplacian on R.

We shall now define a monotonic sequence $M = \{M_k\}$ as follows:

$$M_k = j^2 p_j \text{ for } k = p_{j-1} + j, p_{j-1} + j + 1, \cdots, p_j + j. \tag{4.5}$$

Here $\{p_j\}$ is a sequence with $p_0 = 1$, $p_j > p_{j-1} + j + 2b_j$ for all $j \geqq 1$, and for all $j \geqq 1$ we have $p_j = 2m_{j'}$ for some j'.

First we note that

$$\sum 1/M_k = \sum (1/j^2) \cdot ([p_j + j - (p_{j-1} + j - 1)]/p_j) < \infty.$$

Thus, by Proposition 4.1 there exists a function $g_1 \in A_M$, support of g_1 contained in $[-1 < x < 1]$. Define the function g on R by $g(x) = g_1(x_1) g_2(x_2), \cdots, g_1(x_n)$. I claim that $f \notin g * \mathcal{D}'$.

For, suppose $f = g * W$ for some $W \in \mathcal{D}'$. Since the values of $(g * W)(x)$ for $|x| < 2$ depend only on the values of W on $|x| < 3$, we may assume that W is of the form $\Delta^l h$ for some continuous function h and some l. Then we would have

$$f = g * W = \Delta^l g * h.$$

Moreover, for any r we have

$$\Delta^r f = \Delta^{l+r} g * h. \tag{4.6}$$

We now estimate $\Delta^{l+r_j} g$ by use of the fact that $g_1 \in A_M$. We assume $j > 2l$. First, since $g_1 \in A_M$, we have for any s and any x,

$$\begin{aligned} |(\Delta^s g)(x)| &= |(\partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_n^2)^s g_1(x_1) \cdots g_n(x_n)| \\ &\leqq B^n \sum_{t_1+t_2+\cdots+t_n=2s} |g_1^{(t_1)}(x_1) g_1^{(t_2)}(x_2) \cdots g_1^{(t_n)}(x_n)| \\ &\leqq B^n \sum_{t_1+\cdots+t_2+\cdots+t_n=2s} K^{t_1} M_{t_1}{}^{t_1} \cdots K^{t_n} M_{t_n}{}^{t_n} \\ &\leqq B^n n^{2s} K^{2s} M_{2s}{}^{2s} \end{aligned}$$

because there are n^{2s} terms in the above sum and $\{M_j\}$ is monotonic.

Hence, if $2m_{j'} = p_j$ for some j, we have by (4.5)

$$\begin{aligned} |(\Delta^{l+m_j'} g)(x)| &\leqq B^n (nK)^{2l+2m_j'} M_{2l+2m_j'}{}^{2l+2m_j'} \\ &= B^n (nK)^{2l+2m_j'} (2j^2 m_{j'})^{2l+2m_j'} \\ &= B^n (nKj^2)^{2m_j'} (2m_{j'})^{2m_j'} (nKm_{j'})^{2l}. \end{aligned} \tag{4.7}$$

Now, $(\log x)/x$ is monotonic decreasing for $x \geqq e$, so for $l \geqq 1$ and $m_{j'}$ sufficiently large, $(nKm_{j'})^{2l} \leqq (nK)^{2l} (2l)^{m_j'}$. Thus, (4.7) implies

$$\begin{aligned} |(\Delta^{l+m_j} g)(x)| &\leqq B_1 K_1{}^{2m_j'} (2m_{j'})^{2m_j'} \\ &\leqq B_2 K_2{}^{2m_j'} (2m_{j'})! \end{aligned} \tag{4.8}$$

by Stirling's formula, for certain positive constants B_1, B_2, K_1, K_2.

From (4.6) and (4.8) we deduce that for all x with $|x| \leqq 1$ we have

$$|(\Delta^{m_j'} f)(x)| \leqq B_3 K_2{}^{2m_j'} (2m_{j'})! \tag{4.9}$$

for an infinite number of $m_{j'}$. This clearly contradicts (4.4) and our theorem is proven.

Remark. By a similar kind of argument we can show $(n=1)$ that the intersection of all Carleman non quasi-analytic classes, that is all A_M which are non quasi-analytic, is just A.

We wish now to prove a partial converse of Theorem 4.2. We shall prove also some similar results.

For each $r>0$, denote by A_r the space of functions on R which can be extended to be analytic in the strip C_r defined by $|\mathfrak{I}x|<r$. The topology of A_r is defined by uniform convergence on the compact sets of C_r. This topology is best described by the methods of the theory of infinite derivatives (see [14] for this and following).

By A'_r we denote the dual of A_r; $\hat{A}'_r$ is the Fourier transform of A'_r.

PROPOSITION 4.3. *$\hat{A}'_r$ consists of all entire functions F of exponential type which satisfy*

$$F(z)=O(\exp(l\,|\,\mathfrak{I}z\,|+r'\,|\,\mathfrak{R}z\,|)) \tag{4.10}$$

for some l, r' with $r'<r$. The topology of $\hat{A}'_r$ can be described as follows: Let $H(z)$ be and continuous positive function on C with the property that for any l, r' with $r'<r$ we have $\exp(l\,|\,\mathfrak{I}z\,|+r'\,|\,\mathfrak{R}z\,|)=O(H(z))$. *Call N_H the sets of $F\in\hat{A}'_r$ which satisfy*

$$|F(z)|\leqq H(z) \ \ \textit{for}\ z\in C. \tag{4.11}$$

The sets N_H form a fundamental system of neighborhoods of zero in $\hat{A}'_r$.

Proof. For $n=1$ the fact that $\hat{A}'_r$ consists of all entire functions F of exponential type which satisfy (4.10) is a consequence of Polyà's theorem on conjugate diagrams (see e.g. Boas, *Entire Functions*). The statement about the topology can then be proven by combining Polyà's method with my method of describing the topology of $\boldsymbol{H}'$ (see [7]). The passage to $n>1$ presents no difficulties.

A second proof is as follows: Let $\mathcal{E}(C_r)$ be the space of indefinitely differentiable functions on C_r with the usual topology. The topology of the Fourier transform $\boldsymbol{E}'(C_r)$ of $\mathcal{E}'(C_r)$ can be described by the methods of Part III (see [7]). The passage from $\mathcal{E}'(C_r)$ to $\hat{A}'_r$ is accomplished by means of the fundamental principle for systems of constant coefficient equations (see [15], [16]).

Now, let $S\in\mathcal{E}'$. Then $f\to S*f$ is clearly a continuous linear map of $A_r\to A_r$. We claim this map is *onto*. If we make use of the method outlined in the beginning of Section 2, we have to prove that J cannot be too small at infinity, roughly speaking, we have to show that $J(z)>\text{const}\exp(-\epsilon\,|\,z\,|)$

for "enough" $z \in R$. This inequality is obtained by means of subharmonicity (or rather pleurisubharmonicity) of $\log|J|$. More precisely

PROPOSITION 4.4. *We have*

$$\int_R [\log|J(z)|/(1+|z|^2)]\,dz > -\infty. \tag{4.12}$$

Proof. As in Section 2, we may assume J is bounded on R or even that J is bounded in $\mathcal{I}z_1 \geqq 0, \cdots, \mathcal{I}z_n \geqq 0$. By effecting a translation in z and multiplying J by a suitable constant, we may assume that $J(i, i, \cdots, i) = 1$.

We now prove the proposition by induction on n of the stronger proposition:

$$\int_R [\log|J(z_1, \cdots, z_n)|/(1+|z_1|^2)\cdots(1+|z_n|^2)]\,dz \geqq \log|J(i, i, \cdots, i)|. \tag{4.13}$$

For $n = 1$, inequality (4.13) is a well-known consequence of the subharmonicity of $\log J$. Assume (4.13) is true for values of n smaller than the one in question. Then whenever $\mathcal{I}z_1 \geqq 0$ we have

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[\log|J(z_1, z_2, \cdots, z_n)|/(1+|z_2|^2)\cdots(1+|z_n|^2)]\,dz_2\cdots dz_n \geqq \log|J(z_1, i, \cdots, i)|. \tag{4.14}$$

Now, for fixed $z_2, \cdots, z_n$ the function $z_1 \to \log|J(z_1, z_2, \cdots z_n)|$ is clearly subharmonic and bounded from above. We deduce immediately that the left side of (4.14) is subharmonic and bounded from above for $\mathcal{I}z_1 \geqq 0$. If we call the left side of (4.14) $\tilde{J}(z_1)$, then the well-known results of subharmonicity show that

$$\int_{-\infty}^{\infty}[\tilde{J}(z_1)/(1+|z_1|^2)]\,dz_1 \geqq \tilde{J}(i) \geqq \log|J(i, i, \cdots, i)| \tag{4.15}$$

which is the desired result.

We can now obtain our desired estimates for how rapidly J can decrease at infinity on R:

PROPOSITION 4.5. *Given any $\epsilon > 0$, we can find an A so large that for any $x \in R$ with $|x| > A$ there is a $y \in R$ such that*

$$|x - y| < \epsilon|x| \tag{4.16}$$

and

$$|J(y)| \geqq \exp(-\epsilon|x|). \tag{4.17}$$

Proof. Assume there exists an $\epsilon > 0$ with $\epsilon < \frac{1}{2}$ so that (4.16), (4.17) do not hold for a sequence $\{x^j\}$ with $|x^j| > 2|x^{j-1}|$ and $|x^j| \to \infty$. Let R_j be the ring defined as the set of $x \in R$ with $|x^j| \leqq |x| \leqq (1+\epsilon)|x^j|$. By multiplication by a suitable constant we may assume $|J(x)| \leqq 1$ for all $x \in R$.

Then it is immediate that

$$
\begin{aligned}
&\int_{R_j} [\log|J(z)|/(1+|z_1|^2)\cdots(1+|z_n|^2)]\,dz \\
(4.18)\qquad &\leqq -\epsilon \int_{R_j} [|z|/(1+|z_1|^2)\cdots(1+|z_n|^2)]\,dz \\
&\leqq -k\epsilon\log(1+\epsilon)
\end{aligned}
$$

where $k > 0$ is independent of j. It follows that

$$\Sigma \int_{R_j} [\log|J(z)|/(1+|z_1|^2)\cdots(1+|z_n|^2)]\,dz = -\infty$$

which contradicts Proposition 4.4. This completes the proof of Proposition 4.5.

We can now use the methods of Section 2 (see Lemmas 2.1 and 2.2 and their application in Theorem 2.2) to deduce

THEOREM 4.6. *For any $S \in \mathcal{E}'$ and any $r > 0$, $S * A_r = A_r$.*

In a similar manner we could prove

THEOREM 4.7. *Let $S \in \mathcal{E}'$; let $r > 0$ be fixed. For any $m > 0$ we can find $l, l' > m$ such that if f is analytic in the parallelpiped $|\Re x| < l$, $|\Im x| < r$, then we can find a g which is analytic in $|\Re x| < l'$, $|\Im x| < r$ such that $(S * g)(x) = f(x)$ for $|Rx| < m$, $|\Im x| < r$.*

We wish now to prove the converse of Theorem 4.2, that is, if f is any real analytic function then for any $S \in \mathcal{E}'$ there exists a $g \in \mathcal{E}$ with $S * g = f$. For this purpose, we shall make use of Theorem 4.7 letting m increase to infinity (and $r \to 0$).

Let $S \in \mathcal{E}'$ be fixed; we assume for simplicity of notation that *carrier* $S \subset [-\frac{1}{4} \leqq x \leqq \frac{1}{4}]$. For each m, choose g_m so that g_m is analytic in a parallelpiped and $S * g_m = f$ on $|\Re x| < m+1$, $|\Im x| < r_m$, where r_m is sufficiently small. Now, the functions g_m may not converge on R. However, we can modify g_m by setting $h_m = g_m + k_m$, where $S * k_m = 0$ and k_m is entire. There is now more hope that $\{h_m\}$ will converge.

We have $h_m - h_{m-1} = (g_m - g_{m-1}) = (k_m - k_{m-1})$. Call $k_m - k_{m-1} = l_m$. Then we want to produce l_m so that $S * l_m = 0$ and $(g_m - g_{m-1}) + l_m$ is very

small (and its first m derivatives small) say $\leqq m^{-2}$ on $|\Re x| \leqq m-1$, $\Im x = 0$. We may assume that this inequality holds for smaller values of m. Now, $S*(h_m - h_{m-1}) = 0$ on $|\Re x| \leqq m-1$, $\Im x = 0$. Thus, by a theorem of Malgrange we can choose an exponential polynomial l_m to satisfy the above inequalities. (The result of Malgrange has not been published yet, though it appears in lecture notes. It is closely related to the results of his thesis [19]. I should like to thank Malgrange for pointing this out to me.) The result follows on setting $k_m = k_{m-1} + l_m$.

Now, we can approximate g_m by a polynomial $\tilde{g}_m$ in such a way that $\tilde{g}_m - g_m$ and $S*(\tilde{g}_m - g_m)$ have the property that their first m derivatives are $\leqq m^{-2}$ on $|\Re x| \leqq m$, $\Im x = 0$. This implies that the series

$$\sum [(\tilde{g}_m - \tilde{g}_{m-1}) + l_m]$$

converges in the topology of $\mathcal{E}$, say to g. We have

$$\begin{aligned} S*g &= S*\sum [(\tilde{g}_m - \tilde{g}_{m-1}) + l_m] \\ &= \sum S*[(\tilde{g}_m - \tilde{g}_{m-1}) + l_m] \\ &= \sum S*(\tilde{g}_m - \tilde{g}_{m-1}) \\ &= \lim S*\tilde{g}_m, \end{aligned}$$

where this limit is in the topology of $\mathcal{E}$. Now it is clear that for any $x \in R$,

$$\begin{aligned} \lim (S*\tilde{g}_m)(x) &= \lim (S*g_m)(x) \\ &= f(x). \end{aligned}$$

This, together with Theorem 4.2 gives

THEOREM II. $\bigcap_{S*\mathcal{E}'} S*\mathcal{D}' = \bigcap_{S*\mathcal{E}'} S*\mathcal{E} = A$.

Theorem II is very remarkable. For to show that $\bigcap_{S*\mathcal{E}'} S*\mathcal{D}' \subset$ *real analytic* we used non quasi-analytic classes defined by inequalities of the form $|f^{(r)}(x)| \leqq M_r B^r$. On the other hand, one might suspect that by using non quasi-analytic classes defined by inequalities of the form

$$\left|\sum_{j=1}^{l} f^{(r_j)}(x)\right| \leqq M_{r_1,\cdots,r_l}$$

we could prove a stronger regularity condition on the functions in $\bigcap_{S*\mathcal{E}'} S*\mathcal{D}'$. However, Theorem II shows that no such argument is possible. I do not know if $\bigcap_{S*\mathcal{E}'} S*A = A$. If we would try to prove this by methods similar to our methods of parts I, II, III, or by a method similar to that of Section 2,

we should introduce a "natural" topology in A. Then we should try to describe the Fourier transform of $\hat{A}'$ in a method like that described in the beginning of Section 2. However, it seems extremely unlikely that this is possible for reasons given below:

One possible way of putting a topology on A is to consider A as the union of all spaces of functions analytic in a fixed open set containing R (these space being given the usual compact-open topology) and then giving A the inductive limit topology (see e.g. [14] or [30]). Then we see easily from Proposition 4.3 and the definition of an inductive limit that $\hat{A}'$ consists of all entire functions F of exponential type which satisfy for some l

$$F(z) = O(\exp(l \mid \mathcal{I}z \mid + \epsilon \mid \mathcal{R}z \mid)$$

for all $\epsilon > 0$.

Suppose that the topology of $\hat{A}'$ could be described by means of positive continuous functions $\{H\}$ with the property that a fundamental system of neighborhoods of zero in $\hat{A}'$ consists of those sets N for which there is an H so that N consists of all $F \in \hat{A}'$ which satisfy $\mid F(z) \mid \leqq H(z)$ for all $z \in C$. Now, clearly, for any $F \in \hat{A}'$ and any H there is an $a > 0$ so that $a \mid F(z) \mid \leqq H(z)$ for all $z \in C$. By considering products of entire functions of exponential type zero (constructed by power series) and exponentials we deuce that for each $l > 0$ there is an $\epsilon_l > 0$ so that

$$\exp(l \mid \mathcal{I}z \mid + \epsilon_l \mid \mathcal{R}z \mid) = O(H(z)). \tag{4.19}$$

We shall denote by $\hat{B}'$ the space of functions of $\hat{A}'$ with the topology defined by all functions H which satisfy (4.19).

I claim that inequality (4.19) implies that there exists an ϵ independent of l so that

$$\exp(l \mid \mathcal{I}z \mid + \epsilon \mid \mathcal{R}z \mid) = O(H(z)) \tag{4.20}$$

for all l. This implies immediately that $\hat{B}' \neq \hat{A}'$ and, in fact that a function of B must be analytic in a strip about R. In fact, using Proposition 4.3 we could show that B is the inductive limit of the spaces A_r.

For simplicity I assume $n = 1$. We may clearly assume that the ϵ_l decrease with increasing l. Then for $l > 1$, from the relationships

$$\exp(\mid \mathcal{I}z \mid + \epsilon_1 \mid \mathcal{R}z \mid) = O(H(z))$$
$$\exp(l \mid \mathcal{I}z \mid + \epsilon_l \mid \mathcal{R}z \mid) = O(H(z)).$$

I want to conclude that

$$\exp(l' \mid \mathcal{I}z \mid + \tfrac{1}{4}\epsilon_1 \mid \mathcal{R}z \mid) = O(H(z)),$$

where $l' \to \infty$ with l. Thus I want to show that

$$\begin{aligned}(4.21)\quad &\exp(l' \mid \mathfrak{J}z \mid + \tfrac{1}{4}\epsilon_1 \mid \mathfrak{R}z \mid) \\ &\leqq \text{const.}\max\{\exp(\mid \mathfrak{J}z \mid + \epsilon_1 \mid \mathfrak{R}z \mid), \exp(l \mid \mathfrak{J}z \mid + \epsilon_l \mid \mathfrak{R}z \mid)\}.\end{aligned}$$

To prove (4.21) we may suppose for simplicity that we are in the quadrant $C^{(1)}$: $\mathfrak{J}z \geqq 0$, $\mathfrak{R}z \geqq 0$. Then let $C_1^{(1)}$ be the subset where $l\mathfrak{J}z \geqq \frac{1}{2}\epsilon_1 \mathfrak{R}z$ and let $C_2^{(1)}$ be the complement. In $C_1^{(1)}$ we have

$$\exp(\tfrac{1}{2}l\mathfrak{J}z + \tfrac{1}{4}\epsilon_1\mathfrak{R}z) \leqq \exp(l\mathfrak{J}z).$$

In $C_2^{(1)}$ we have

$$\exp(\tfrac{1}{2}l\mathfrak{J}z + \tfrac{1}{4}\epsilon_1\mathfrak{R}z) \leqq \exp(\epsilon_1\mathfrak{R}z).$$

Thus (4.21) is proven with $l' = \frac{1}{2}l$ is the desired result.

There is another possible method of introducing a topology on A: By our above, A is the projective limit of the spaces $\mathcal{E}_M$, that is, intersection which are Carleman non quasi-analytic. (Here $\mathcal{E}_M$ is the space of functions in E which satisfy inequalities of the form (4.1) on every compact set; $\mathcal{E}_M$ is given a natural topology as in [14].) We could give A the projective limit topology, that is, a convex set $N \subset A$ is a neighborhood of zero if it is the intersection with A of a neighborhood of zero in some E_M. Call K the set of functions of A with this topology. Then we define, as usual, the Fourier transform $\hat{K}'$ of the dual K' of K. Assume the topology of $\hat{K}'$ can be described by functions $\{H\}$ as above for $\hat{B}'$. Then I want to show that these functions H also satisfy (4.20) for some ϵ independent of l.

To prove my assertion, I know (see e.g. [14]) that the set $\{a_j M_j^{-j} z^j\}$ is bounded in B' whenever $a_j = O(\epsilon^j)$ for all $\epsilon > 0$. (For the linear functions $f \to i^j a_j M_j^{-j} f^{(j)}(0)$ form a bounded set in E_M and $a_j M_j^{-j} z^j$ are their Fourier transforms.) Suppose for example that for no $\epsilon > 0$ is $\exp(\epsilon \mid \mathfrak{R}z \mid) = O(H(z))$. Then there exists an infinite sequence of points $\{z_k\}$ with $\{\mid \mathfrak{R}z_k \mid\}$, $\{\mid z_k \mid\}$ sufficiently lacunary (see below) such that $H(z_k) < \exp(1/k^3 \mid \mathfrak{R}z_k \mid)$.

I want to construct $\{M_j\}$ so that for a suitable x_k, we have

$$(4.22)\qquad k^{-j} M_j^{-j} \mid z_k \mid^j = \exp[k^{-3} \mid \mathfrak{R}z_k \mid).$$

For this we need

$$(4.23)\qquad M_j = k^{-1} \mid z_k \mid exp(-j^{-1}k^{-3} \mid \mathfrak{R}z_k \mid).$$

For given k, choose $j = [\mid \mathfrak{R}z_k \mid k^{-3}] + 1$ and for this choice of j define M_j by (4.23). We assume $\{\mid \mathfrak{R}z_k \mid\}$ is lacunary enough so that

$$j(k) = [\mid \mathfrak{R}z_k \mid k^{-3}] + 1$$

is a strictly increasing function of k. The definition of $\{M_j\}$ is completed by setting $M_{j'} = M_j$ whenever $j' < j$, j of the form $[|\mathcal{R} z_k| k^{-3}] + 1$ and for no other j'' with $j' < j'' < j$ is j'' of the form $[|\mathcal{R} z_k| k^{-3}] + \cdot$.

I have only to show that, with this choice of $\{M_j\}$, $\sum M_j^{-1} < \infty$. We have

$$\begin{aligned}\sum M_j^{-1} &\leqq \sum_k k\,|z_k|^{-1}\cdot(j(k) - j(k-1))\\ &\leqq \sum_k k\,|z_k|^{-1} j(k)\\ &\leqq \sum_k k\,|z_k|^{-1}\{[|\mathcal{R} z_k|\,k^{-3}] + 1\}\\ &\leqq \sum k\,|z_k|^{-1} + \sum k^{-2}\\ &< \infty\end{aligned}$$

if we assume, as we may, that $|z_k| \geqq k^3$.

This completes the proof of our assertion that for some ϵ_0 we must have $\exp(\epsilon_0 |\mathcal{R} z|) = O(H(z))$. (The above shows even that $\exp(\epsilon_0 |z|) = O(H(z))$.) By considering products of functions of the form $a_j M_j^{-j} z_j^k$ with exponentials we may deduce by the above method that, for any H as above and any $l > 0$, we can find an ϵ_l so that $\exp(\epsilon_l |\mathcal{R} z| + l |\mathcal{I} z|) = O(H(z))$. Thus, as in the example of $\hat{B}'$ above, the functions H are not sufficient to define the topology of $\hat{K}'$.

Remark. I do not know if the topologies of K and A are the same.

5. Elliptic operators. In this section we shall consider C^∞ and entire elliptic operators $S \in \mathcal{E}'$, and we shall characterize them completely. In order to explain the principles which underline our theory, we shall first give a heuristic argument in case $n = 1$.

Call J the Fourier transform of S; suppose for simplicity that S is invertible. Then we know by the results of Section 3 that each distribution $V \in \mathcal{D}'$ which satisfies $S * V = 0$ can be expanded in a convergent series of exponential polynomial solutions; these latter correspond to the zeros of $J(z)$. When must such a convergent series be C^∞ (entire)?

Let a, b be real numbers and consider $\exp(iax + bx)$. Its derivative is $(ia + b)\exp(iax + bx)$. We have

$$(5.1) \qquad |(d/dx)\exp(iax + bx)| = |ai + b|\exp(bx).$$

The above (5.1) shows that for any $l > 0$,

$$(5.2) \qquad \max_{|x|\leqq l} |(d/dx)\exp(iax + bx)| \leqq \max_{|x|\leqq l + |\log|ai+b||/|b|} \exp(bx) = \max_{|x|\leqq l + |\log|ai+b||/|b|} \exp(iax + bx).$$

From (5.2) it follows that if we have a series $\sum C_j \exp(ia_j x + b_j x)$ which converges absolutely and uniformly on every compact set, say to $f(x)$, then we can obtain bounds for f' on an interval $|x| \leqq l$ in terms of bounds for f on some interval $|x| \leqq l'$ if we know that we can find an $M > 0$ so that, for all j,

$$(5.3) \qquad |\log|a_j i + b_j||/|b_j| \leqq M.$$

Hence, by repeating this argument and noting that any finite number of terms of the series $\sum c_j \exp(ia_j x + b_j x)$ do not affect the question of whether $f \in \mathcal{E}$, we see that a sufficient condition to know that $f \in \mathcal{E}$ is

$$(5.4) \qquad \limsup |\log|a_j i + b_j||/|b_j| < \infty.$$

By a similar argument, we would know that f is entire if

$$(5.5) \qquad \limsup |a_j|/|b_j| < \infty.$$

We are therefore led to guess that (5.4) (where $a_j - ib_j$ are the zeros of J) is the condition that S be weakly C^∞ elliptic, and (5.5) is the condition that S be entire elliptic. We shall see that this is actually the case.

The proof of one half of the above, namely, if S does not satisfy condition (5.4), or (5.5), then S is not weakly C^∞ elliptic, or entire elliptic in x_1, will be accomplished by means of exponential sums, that is, e.g., if S does not satisfy (5.5), then we can find a series $\sum c_j \exp(iz^j \cdot)$, where $J(z^j) = 0$, which converges in the space $\mathcal{E}$ to a function which is not analytic in x_1.

In the theory of elliptic differential equations there are in the literature essentially two different methods to obtain results of the type of Theorem III. The first method depends essentially on Gärding's lemma (see [17]) which in turn depends on the fact that J is large at infinity on R. Since, as we shall see later, there exist $S \in \mathcal{E}$ which are C^∞ elliptic but are small at infinity, it does not seem that such a method can be of use in our case.

The second method depends upon the construction of an elementary solution for S which is a C^∞ (or analytic) function outside of a neighborhood of the origin. This method cannot hope to succeed in case S is not invertible for then by Theorem 2.2 there exists no elementary solution which is a distribution in the sense of Schwartz. In case S is invertible, this method does work, and it is outlined in my note [10], and will be presented in detail below. However, even in case S is not invertible, there is hope to find an elementary solution with desired properties which is not a distribution in the sense of Schwartz but which is a continuous linear function on a suitable space of non quasi-analytic functions. However, as Theorem 6.2 shows, this method cannot work for arbitrary S.

There is a third possibility for proving our results. Let us reexamine the case $n=1$. It would be fairly easy to obtain our above heuristic argument with the methods of Section 3 to obtain the desired result in case S is invertible. Even in case S is not invertible by combining these methods with certain properties of Abel limits, there seems to be hops to prove an extension of our results, but we shall not discuss this here. This method is fairly close to my fundamental principle and will be discussed elsewhere (see [16]).

Before proving our assertions about ellipticity, I wish to state several preliminary propositions on distributions which are C^∞ in x_1:

PROPOSITION 5.1. *Let $T \in D'$ be C^∞ in $(x_1, \cdots, x_r)$; then for $h(x_{r+1}, \cdots, x_n)$ an indefinitely function of compact support, we have $T * h \in \mathcal{E}$, that is, T is C^∞ in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$.*

(Here by $T * h$ we mean the convolution of T with the direct product (see [24]) of h with the δ distribution in $(x_1, \cdots, x_r)$.)

Proof. We prove the result in case $r=1$ as the general proof is similar. For any $j_1, j_2, \cdots, j_n$, we have

$$(5.6)\qquad \begin{aligned}&(\partial^{j_1+\cdots+j_n}/\partial x_1^{j_1}\cdots\partial x_n^{j_n})(T*h)\\ &\qquad = (\partial^{j_1}/\partial x_1^{j_1})T * (\partial^{j_2+\cdots+j_n}/\partial x_2^{j_2}\cdots\partial x_n^{j_n})h.\end{aligned}$$

Now, let K be any compact set in R. Then because h is of compact support, we can find a compact set $L \subset R$ so that the values of $T * h$ on K depend only on the values of T on L. By definition, the distributions $(\partial^{j_1}/\partial x_1^{j_1})T$ are of bounded order on L. Thus, the right side of (5.6) is of bounded order on K (this bound is independent of $j_1, j_2, \cdots, j_n$). Hence, al lthe derivatives of $T * h$ are of bounded order on K, which proves (see [24]) that $T * h$ is an indefinitely differentiable function on K. Since K was arbitrary, it follows that $T * h \in \mathcal{E}$. Thus Proposition 5.1 is proven.

Note that a similar argument shows that if T is entire in x_1, then $T * h$ is entire in x_1.

For T a distribution which is C^∞ in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$ we define the restriction $T_{x_1=0,\cdots,x_r=0}$ of T to the plane $x_1 = x_2 = \cdots = x_r = 0$ by

$$(5.7)\qquad T_{x_1=0,\cdots,x_r=0}\cdot h = (T * h)(0)$$

for any $h \in \mathcal{D}(x_{r+1}, \cdots, x_n)$. We prove that this restriction is a distribution. More generally, we have

PROPOSITION 5.2. *If T is C^∞ in $(x_1, \cdots, x_r)$* relative to $(x_{r+1}, \cdots, x_n)$,

then $h \to T * h$ is a continuous map of $\mathcal{D}(x_{r+1}, \cdots, x_n)$ into $\mathcal{E}$. A necessary and sufficient condition for $W \in D'$ to be C^∞ in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$ is that the orders of $\{(\partial^j/\partial x_1^{j_1} \cdots \partial x_r^{j_r})W\}$ should be zero in $x_1, \cdots, x_r$, that is, for any $j_1, \cdots, j_r$ and for $K \subset R$ compact, we can find a differential operator D in $x_{r+1}, \cdots, x_n$ and a measure μ on K so that

$$(\partial^j/\partial x_1^{j_1} \cdots \partial x_r^{j_r})W = D\mu. \tag{5.8}$$

Proof. Let K' be a fixed compact set in $(x_{r+1}, \cdots, x_n)$. Then $h \to T * h$ is a linear map of $\mathcal{D}_{K'}(x_{r+1}, \cdots, x_n) \to \mathcal{E}$. Moreover, this map is closed, for if $h \to h'$, then $T * h \to T * h'$ in the topology of $\mathcal{D}'$, so if $T * h$ converges in the topology of $\mathcal{E}$, it can only converge to $T * h'$. Thus by the closed graph theorem, $h \to T * h$ is continuous on each $\mathcal{D}_{K'}(x_{r+1}, \cdots, x_n)$. Hence, the map is continuous on $\mathcal{D}(x_{r+1}, \cdots, x_n)$ by the definition of an inductive limit.

Next, if W satisfies the condition stated, then arguing as in the proof of Proposition 5.1 we see that W is C^∞ in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$. Conversely, let W be C^∞ in $(x_1, \cdots, x_r)$ relative to $(x_{r+1}, \cdots, x_n)$. For simplicity of notation we assume $n = 2$, $r = 1$; the general case is treated similarly. Suppose there is a cube K such that the derivatives $\{(\partial^j/\partial x_1^j)W\}$ are not of zero order in x_1 on K. Then there exists a j and sequence of functions f^k with supports contained in K so that

$$\max |(\partial^j/\partial x_1^j)(\partial^l/\partial x_2^l)(f^k)(x)| \leqq 1 \quad \text{for } l = 0, 1, \cdots, j$$

but

$$|W \cdot (\partial^j/\partial x_1^j)f^k| \geqq k.$$

Now, the set $\{(\partial^j/\partial x_1^j)f^k_{x_1=a}\}$ is clearly bounded in $\mathcal{D}$. ($f^k_{x_1=a}$ is the function $x_2 \to f^k(a, x_2)$.) Thus, by the first part of Proposition 5.2, the set $\{W * (\partial^j/\partial x_1^j)f^k_{x_1=a}\}$ is bounded in $\mathcal{E}$. Since $\{a\}$ is compact and since $(\partial^j/\partial x_1^j)f^k_{x_1=a}$ and hence $W * (\partial^j/\partial x_1^j)f^k_{x_1=a}$ depend continuously on a, it follows that $\{\int W * (\partial^j/\partial x_1^j)f^k_{x_1=a}\, da\}$ is bounded. But,

$$|\int W * (\partial^j/\partial x_1^j)f^k_{x_1=a}\, da| = |W \cdot (\partial^j/\partial x_1^j)f^k| \to \infty$$

which completes the proof of Proposition 5.2.

PROPOSITION 5.3. *Suppose J does not satisfy (1.1) and (1.2), that is, there exists a sequence of points $\{{}_jz\}$ and a $k > 0$ such that $J({}_jz) = 0$, $|{}_jz| \to \infty$, $|{}_jz_1|^k \geqq |{}_jz|$ for j large enough, but*

$$\limsup |\mathfrak{I}({}_jz)|/\log(1 + |{}_jz_1|) = M < \infty.$$

Then S is not C^∞ elliptic in x_1; in fact, there exists a $T \in \mathcal{D}'$ with $S * T = 0$, and T not C^∞ in x_1. Given any q, we may even choose T to be a q-times differentiable function on $|x| \leqq q$.

Proof. I shall assume that the sequence $\{|z^j|\}$ is strictly increasing to infinity and is, in fact, sufficiently lacunary to satisfy the conditions below (all this may be assured by taking a suitable subsequence). We may also assume for simplicity that $\mathfrak{I}({}_jz_a) \geqq 0$ for all j, a. We can choose the ${}_jz$ so lacunary that $|{}_{j+1}z| \geqq |{}_jz| + j$. Call $W = \sum \delta_{jz}$; it is clear from our explicit expression for the topology of $\boldsymbol{D}$ that the series $\sum \delta_{jz}$ converges in $\boldsymbol{D}'$. Thus, the Fourier transform T of W lies in $\mathfrak{D}'$ and satisfies $S * T = 0$.

I claim that for no sequence $\{c_l\}$ of positive numbers is $\{c_l(\partial^l/\partial x_1{}^l)T\}$ bounded in $\mathfrak{D}'$; that is, T is not C^∞ in x_1. For, assume that for some $\{c^l\}$ it is true that $B = \{c_l(\partial^l/\partial x_1{}^l)T\}$ is bounded in $\mathfrak{D}'$. Then it follows from the definition of the topology of $\mathfrak{D}'$ that we can find an r so large that B is bounded on the bounded sets of $\mathfrak{D}_1{}^r$ (r-times differentiable functions with supports in $|x| \leqq 1$ with Schwartz [24] topology). I shall show that this is impossible.

Using the methods of the last part of the proof of Theorem 2.2 we can construct a sequence of function $\{F_t\} \subset \mathfrak{D}_1$ such that

1. $F_t(0) = 1$,
2. $|F_t(z)| \leqq 1$ for $\mathfrak{I}z_1 \leqq 0, \mathfrak{I}z_2 \leqq 0, \cdots, \mathfrak{I}z_n \leqq 0$,
3. $|F_t(z)| \leqq (t/z)^t \exp|\mathfrak{I}z|$ for all z.

Define

$$G_t(z) = F_t(z - z_t)/|{}_tz|^{r+2}. \tag{5.9}$$

Then we see immediately that

4. $\{G_t\}$ is bounded in $\mathfrak{D}_1{}^r$,
5. For any s we can find t_0 so large that

$$|z_1{}^sW \cdot G_t| = |\sum_j {}_jz_1{}^s G_t({}_jz)| \geqq \tfrac{1}{2}|{}_tz_1{}^s|/|{}_tz|^{r+2} \tag{5.10}$$

for $t \geqq t_0$. Hence, if s is large enough, it follows that $\{|z_1{}^sW \cdot G_t|\}$ is not bounded.

We have proved that for any r we can find an s large enough so that $z_1{}^sW$ is not bounded on the bounded sets of $\mathfrak{D}_1{}^r$, hence, $(\partial^s/\partial x_1{}^s)T$ is not bounded on the bounded sets of $\mathfrak{D}_1{}^r$. Hence, T is not C^∞ in x_1 so our result is established.

By considering sums of the form $\sum |{}_jz|^{-p}\delta_{jz}$ we would, given q, produce a distribution T which satisfies $S * T = 0$, is C^q on the set $|x| \leqq q$ and is not C^∞ in x_1 by taking p large enough. (However, we cannot, in general, produce a T which is a distribution of finite order as follows from results below.)

By reasoning as in Proposition 5.3 we can establish

PROPOSITION 5.4. *Suppose there exists a sequence of points* $\{{}_jz\}$ *and a* $k>0$ *such that* $J({}_jz)=0$, $|{}_jz|\to\infty$, $|{}_jz_1|^k \geqq |{}_jz|$ *for* j *large enough, but*

$$\limsup |\Im({}_jz)|/\log(1+|{}_jz_1|)=0. \tag{5.11}$$

Then S *is not weakly* C^∞ *elliptic in* x_1. *Given any* q, *we can find a function* f *which is* q *times differentiable and satisfies* $S*f=0$ *but which is not* C^∞ *in* x_1 *(even considered as an element of* $\mathcal{D}'_F$*).*

We now consider the analog for analytic elliptic:

PROPOSITION 5.5. *Suppose there exists a sequence of points* $\{{}_jz\}$ *and a* $k>0$ *such that* $J({}_jz)=0$, $|{}_jz_1|\to\infty$, $|{}_jz_1|>k\log|{}_jz|$ *for* j *large enough, but*

$$\limsup |\Im({}_jz)|/|{}_jz_1|=0. \tag{5.12}$$

Then J *is not entire elliptic in* x_1. *We can find a function* $f\in\mathcal{E}$ *with* $S*f=0$ *such that* f *is not entire in* x_1.

Proof. We may clearly assume that $|{}_{j+1}z|\geqq|{}_jz|+j$, and, as in the proof of Proposition 5.1, that $\Im({}_jz_a)\geqq 0$ for all j, a. Now the series $\sum\delta_{jz}$ may not converge in $\mathcal{D}'$ if $\Im({}_jz)$ is too large. However, if $\{b_j\}$ is any sequence of positive numbers increasing to infinity, then the series

$$\sum \exp(-b_j|\Im({}_jz)|)\delta_{jz}$$

converges in $\mathcal{D}'$ as is readily verified. We shall choose

$$b_j=|{}_jz_1|^{\frac12}(|\Im({}_jz)|+1)^{-\frac12}. \tag{5.13}$$

By our hypothesis, $b_j\to\infty$. Next, let $c_j\to\infty$ at a very slow rate (to be specified later). It is readily verified that the series

$$W=\sum|{}_jz|^{-c_j}\exp(-b_j|\Im({}_jz)|)\delta_{jz} \tag{5.14}$$

converges in the topology of $\mathcal{E}$. If f denotes the Fourier transform of W, then $f\in\mathcal{E}$ satisfies $S*f=0$. I claim that if the $|{}_jz|$ are lacunary enough (to be explained later), then f is not entire in x_1, even if f is considered as an element of D'.

If f were entire in x_1, then for r large enough

$$B=\{(1/s!)(\partial^s/\partial x_1^s)f\}$$

would be bounded on the bounded sets of $\mathcal{D}_1^r$. We shall show that this is impossible if $c_j\to\infty$ slowly enough and if the numbers $|{}_jz|\to\infty$ fast enough.

Let us note the following: If $s = [|z|/2]$, then

$$
\begin{aligned}
|z|^s/s! &\geqq \exp([|z|/2]\log|z| - [|z|/2]\log[|z|/2]) \\
&\geqq \exp(2[|z|/2]). \\
&\geqq e^{-1}\exp(|z|).
\end{aligned} \tag{5.15}
$$

We assume the points ${}_jz$ are so lacunary that $|{}_{j+1}z| \geqq e\,|{}_jz|$. We now return to the notation of the proof of Proposition 5.3. The functions G_t satisfy 4. Moreover, if $s = [|{}_tz_1|/2]$, then by (5.15) and the definition of b_t, we have

$$
\begin{aligned}
|(1/s!)\,{}_tz_1{}^s|\exp(-b_t|\mathfrak{I}({}_tz)|) &\geqq \exp(-1 + |{}_tz_1| - |{}_tz_1|^{\frac{1}{2}}|\mathfrak{I}({}_tz)|^{\frac{1}{2}}) \\
&\geqq \exp(\tfrac{1}{2}|{}_tz_1|)
\end{aligned} \tag{5.16}
$$

for t large enough because of (5.12). Thus, making use of the properties of the G_t, we have, for this s if t is large enough,

$$
\begin{aligned}
&|(1/s!)z_1{}^s W\cdot G_t| \\
&= |\sum_j (1/s!)\,{}_jz_1{}^s G_t({}_jz)\,|{}_jz|^{-c_j-r-2}\exp(-b_j|\mathfrak{I}({}_jz)|)| \\
&\geqq \tfrac{1}{2}\exp(\tfrac{1}{2}|{}_tz_1|)\,|{}_tz|^{-c_j-r-2}.
\end{aligned} \tag{5.17}
$$

We can clearly choose $\{c_j\}$ with $c_j \to \infty$ so that for any r the right hand side is unbounded. This completes the proof of Proposition 5.5.

Remark. The same method can be used to show that f is not real analytic in x_1 in the neighborhood of any point.

We are now ready to prove the converses of these three propositions.

Let $T \in \mathfrak{D}'$; we say T is C^∞ in x_1 on an open set $\Omega \subset R$ if we can find positive numbers b_j so that $\{b_j(\partial^j/\partial x_1{}^j)T\}$ is a bounded set of distributions on Ω. (A similar definition for T entire in x_1 on Ω.) Let $S \in \mathcal{E}'$; we say that $P \in \mathfrak{D}'$ is a C^∞ *in* x_1 *parametrix for* S if P is C^∞ in x_1 outside of some neighborhood of the origin and

$$S * P = \delta + W, \tag{5.18}$$

where W is C^∞ in x_1. We say that P is a C^∞ *in* x_1 *parametrix of finite order* if $W, P \in \mathfrak{D}'_F$ and if we can find an m so that for each $r > 0$ we can find an e_r so that the first r derivatives of P with respect to x_1 are distributions of order $\leqq m$ for $|x| > e_r$. We say that P is an *entire in* x_1 *parametrix* for S if W is analytic in x_1 for $|\mathfrak{I}x_1| < r$ and if given any $r > 0$ we can find a d_r so that P is analytic in x_1 for $|\mathfrak{I}x_1| < r$ in $|x| > d_r$.

The importance of the parametrix comes from

PROPOSITION 5.6. (a) *If $S \in \mathcal{E}'$ has a C^∞ in x_1 parametrix, then S is C^∞ elliptic in x_1.*

(b) *If $S \in \mathcal{E}'$ has a C^∞ in x_1 paramerix of finite order, then S is weakly C^∞ elliptic in x_1.*

(c) *If $S \in \mathcal{E}'$ has an entire in x_1 parametrix, then S is entire elliptic in x_1.*

Proof. All the proofs are similar, so we prove (c) for illustration. Let $S * T$ be entire in x_1; we have to show that T is entire in x_1. We show that, in the neighborhood of the origin $|x| < m$, T is analytic in x_1 for $|\mathfrak{I} x_1| < r$. Let $h \in D$ be 1 for $|x| < r$. Let $h \in D$ be 1 for $|x| < m'$ (to be chosen later). Then consider hT. We have by (5.18)

$$
\begin{aligned}
hT &= hT * \delta \\
&= hT * P * S - hT * W.
\end{aligned}
\tag{5.19}
$$

Now, $hT * W$ is certainly entire in x_1 since W is. Moreover, $h * P * S = (S * hT) * P$. We know that if m' is large enough, $S * hT = S * T$ on $|x| < m''$ (where m'' can be made arbitrarily large). Now, the restriction of $(S * hT) * P$ to $|x| < m$ depends on the convolution of P with the restriction of $S * hT$ to the set $|x| < m'$ (m' large enough) and on the convolution of hT with the restriction of P to $|x| > m'''$ (which again can be made arbitrarily large). Thus, our assertion is established.

PROPOSITION 5.7. *Let $S \in \mathcal{E}'$ be invertible for $\mathfrak{D}'$ and satisfy* (1.1) *and* (1.2). *Then there exists a C^∞ in x_1 parametrix for S.*

Proof. In order to make the proof clear, we shall give the proof first in case $n = 1$. Then there are only a finite number of real zeros of J, and the modulus of the imaginary parts of the zeros $\rightarrow \infty$ faster than any multiple of the log of the modulus of the real part, We want to define something like

$$
\int_{-\infty}^{\infty} [\exp(ixz)/J(z)]\, dx. \tag{5.20}
$$

Let K be chosen so large that all the zeros z_j of J satisfy $|\mathfrak{R} z_j| < K - 1$. Then instead of the above integral, we shall consider

$$
\int_{-\infty}^{-K} [\exp(ixz)/J(z)]\, dz + \int_{K}^{\infty} [\exp(ixz)/J(z)]\, dz \tag{5.21}
$$

which differs from (5.20) by an integral over a compact set. We shall see that the integral over the compact set does not matter.

Next, we consider $\int_K^\infty [\exp(ixz)/J(z)]\,dz$. Assume x is large and positive. Then it would be natural to try to shift the contour as high as we can in the complex z plane, i.e., make $\mathfrak{I}z$ as large as possible, and in this way make use of the smallness of $\exp(ixz)$. This can be done because J has no zeros unless $\mathfrak{I}z$ is very large.

The only trouble we encounter is that $\int_K^\infty [\exp(ixz)/J(z)]\,dz$ does not have to have a meaning in the usual sense because J may $\to\infty$ at infinity. Therefore we modify thet argument slightly as follows: Let $f \in \mathfrak{D}$; then instead of (5.21) we consider

$$\int_{-\infty}^{-K} [F(z)/J(z)]\,dz + \int_K^\infty [F(z)/J(z)]\,dz. \tag{5.22}$$

We shall show that each integral on the right of (5.22) exists in the usual sense (because J is slowly decreasing) and if f vanishes for $x < a$, where $a > 0$, will be prescribed later, then we can shift the contour as we described above. We then define $P \in \mathfrak{D}'$ by

$$P \cdot g = \int_{-\infty}^{-K} [G(z)/J(z)]\,dz + \int_K^\infty [G(z)/J(z)]\,dz \tag{5.23}$$

for any $g \in D$. We verify easily that P is a parametrix for S. Moreover, the above argument that P is C^∞ for $x > a$ (similarly, for $x < -a$). Thus our result will be proven.

Now we proceed with the details of the proof: Since J is slowly decreasing, there exists an $A > 0$ so that for each $z \in R$ we can find a $w \in R$ $(w = w(z))$ with $|w - z| < A \log(1 + |z|)$ and $|J(w)| \geqq (A + |z|)^{-A}$. If $|z|$ is large enough (and z_0 real), then the circle $|z - z_0| \leqq 2A \log(1 + |z_0|)$ will contain no zeros of J. Now by the minimum modulus theorem (see [8]) we can draw about $w_0 = w(z_0)$ a circle in the complex z plane of radius between $(3/2)A \log(1 + |z_0|)$ and $2A \log(1 + |z_0|)$ on which $|J(z)| \geqq (A' + |z_0|)^{-A'}$ for some A'. But $1/J$ is analytic in this circle. Thus we also have $|J(z_0)| \geqq (A' + |z_0|)^{-A'}$ by the maximum modulus theorem. Hence, for any $F \in \boldsymbol{D}$ the right side of (5.23) exists as an absolutely convergent integral and P defined by (5.23) is a distribution (of finite order).

Next we want to shift the contour in $\int_K^\infty [F(z)/J(z)]\,dz$ if f vanishes for $x < a$. We pick a sequence of numbers K_j with $K_0 = K$, $K_{j+1} \geqq K_j + 1$ and so that for z' real, $z' \geqq K_j$, $J(z)$ does not vanish if $|z - z'| \leqq 2j \log(1 + |z'|)$.

We define a curve Γ by: Γ consists of all z for which $\mathfrak{I}z = j \log Rz$ if

$K_j \leqq Rz \leqq K_{j+1}$, and Γ is made continuous by joining the various parts by vertical lines. It is trivial to verify that the total length of Γ lying above the interval $[K_0, b]$ is $<$ const. b^2. Moreover, we can again apply our minimum modulus-maximum modulus argument as above to deduce that

$$(5.24) \qquad |J(z)| \geqq (A'' + |z|)^{-A''} \exp(-A'' |\vartheta z|)$$

for $z \in \Gamma$. On the other hand, f vanishes for $x \leqq a$; a simple argument shows that this implies that

$$(5.25) \qquad |F(z)| \leqq A'''(A'' + |z|)^{-A''-4} \exp(-(A''+1)|\vartheta z|)$$

if a is large enough. Moreover, if $f \in B$, where $B \subset \mathfrak{D}$ is a bounded set of functions which vanish for $x < a$, then we may assume that inequality (5.25) holds for all $f \in B$ for the same A'''.

All the above shows us that $\int_\Gamma [F(z)/J(z)]dz$ exists in the absolute sense. Moreover, again using the fact that f vanishes for $x < a$, we deduce immediately that

$$(5.26) \qquad \int_K^\infty [F(z)/J(z)]dz = \int_\Gamma [F(z)/J(z)]dz.$$

A similar construction holds for $\int_{-\infty}^{-K} [F(z)/J(z)]dz$.

We now want to show that P is C^∞ for $x > a$. (A similar method works for $x < -a$.) Let B be a bounded set in $\mathfrak{D}$ of functions which vanish for $x > a$. Then for $f \in B$ and for any r we deduce as above

$$(5.27) \qquad \int_K^\infty z^r[F(z)/J(z)]dz = \int_\Gamma z^r[F(z)/J(z)]dz.$$

By the construction of Γ, we have on Γ,

$$(5.28) \qquad |z^r| \exp(-|\vartheta z|) \leqq b_r.$$

Combining this with (5.24) and (5.25) we deduce that for all $f \in B$,

$$(5.29) \qquad \left| \int_K^\infty z^r[F(z)/J(z)]dz \right| \leqq b_r A''''.$$

Here b_r depends only on r and J (but is indepenednt of B) and A'''' depends only on B.

A similar construction for $\int_{-\infty}^{-K}$ shows that

$$(5.30) \qquad |P \cdot f^{(r)}| \leqq 2A''''b_r$$

for all $f \in B$. This means that $\{b_r^{-1}P^{(r)}\}$ is bounded on every bounded set of functions in $\mathfrak{D}$ which vanish for $x < a$, so that $\{b_r^{-1}P^{(r)}\}$ is bounded in $\mathfrak{D}'$ for $x > a$. This means that P is C^∞ for $x > a$; a similar method applies for $x < -a$.

Finally, we must prove that P is a parametrix for S, that is, $S * P = \delta + W$, where W is a C^∞ function. We have for $f \in \mathfrak{D}$,

$$\begin{aligned} S * P \cdot f &= P \cdot S * f \\ &= \int_{-\infty}^{-K} [J(z)F(z)/J(z)]\,dz + \int_K^\infty [J(z)F(z)/J(z)]\,dz \\ &= \int_{-\infty}^{-K} + \int_K^\infty F(z)\,dz \\ &= \int_{-\infty}^{\infty} F(z)\,dz - \int_{-K}^{K} F(z)\,dz \\ &= \delta \cdot f + W \cdot f, \end{aligned}$$

where $W(x) = \int_{-1}^{1} \exp(ixz)\,dz$ is an entire function of exponential type. Thus our result is proven in case $n = 1$.

We show now how to modify the above argument in case $n \neq 1$. Our first task is to define P. Since J is slowly decreasing, by Lemma 2.2 each $z \in R$ can be surrounded by a set on which $|J| \geqq (A + |z|)^{-A}$. Moreover, the maximum distance from z to any point on this set is $\leqq A \log(1 + |z|)$. Call R' the set of $z \in R$ for which $|J(z)| \geqq (A + |z|)^{-A}$, and set $R'' = R - R'$. Then we define P by

$$P \cdot f = \int_{R'} [F(z)/J(z)]\,dz.$$

Then P clearly is a distribution.

Let $z \in R''$, then the above construction shows there must be a point $w \in V$ with $|w - z| < A \log(1 + |z|)$. In particular, $|\mathfrak{J}w| < A \log(1 + |z|)$ and even

$$|\mathfrak{J}w| < A' \log(1 + |w|). \tag{5.31}$$

Thus, we know from the fact that J satisfies (1.1) and (1.2) that for all such z,

$$\liminf \log |w_1| / \log |w| = 0. \tag{5.32}$$

It follows from (5.32) and the fact that $|z - w| < A \log(1 + |z|)$ that we also have

$$\liminf_{z \in R'', |z| \to \infty} \log |z_1| / \log |z| = 0. \tag{5.33}$$

This inequality will be used to show that P is a C^∞ in z_1 parametrix for S.

Our next task is to shift the contour from R' to show that P is C^∞ in x_1 outside of a neighborhood of the origin. We show first that P is C^∞ in x_1 in the half-space $x_n > a$ for a sufficiently large. A similar argument applies to $x_n < -a$ and also to the half-spaces $x_j > a$, $x_j < -a$ so that we shall know that P is C^∞ in x_1 for $|x| > a$.

Let $f \in \mathcal{D}$ vanish for $x_n \leqq a$. Let $z_1, \cdots, z_{n-1}$ be fixed. Then we shall define a new contour $\Gamma(z_1, \cdots, z_{n-1})$ in such a way that

$$(5.34) \qquad P \cdot f = \int \cdots \int dz_1 \cdots dz_{n-1} \int_{\Gamma(z_1, \cdots, z_{n-1})} [F(z)/J(z)]\, dz_n.$$

Call $J_{z_1, \cdots, z_{n-1}}$ the function $z \to J(z_1, \cdots, z_{n-1}, z_n)$. For fixed $z_1, \cdots, z_{n-1}$ we divide the real z_n line into subsets $K_j(z_1, \cdots, z_{n-1})$. $K_j(z_1, \cdots, z_{n-1})$ consists of all real z_n with the property that $J_{z_1, \cdots, z_{n-1}}(z'_n)$ does not vanish for $|z'_n - z_n| \leqq 2j \log(1 + |z_n|)$.

Now we define the $\Gamma(z_1, \cdots, z_{n-1})$ by: For each integer l, $\Gamma(z_1, \cdots, z_{-n1})$ consists (above $[l \leqq t \leqq l+1]$) of those points z_n with $\mathfrak{J} z_n = j \log(1 + |Rz_n|)$, $l \leqq Rz_n \leqq l+1$, if j is the smallest integer such that all points in $[l \leqq t \leqq l+1]$ belong to $R'_{z_1, \cdots, z_{n-1}} \cap K_j(z_1, \cdots, z_{n-1})$. If there is a t in $[l \leqq t \leqq l+1]$ for which $t \notin R'_{z_1, \cdots, z_{n+1}}$, then $\Gamma(z_1, \cdots, z_{n-1})$ above $[l \leqq t \leqq l+1]$ is just $R'_{z_1, \cdots, z_{n-1}} \cap [l \leqq t \leqq l+1]$. Finally, $\Gamma(z_1, \cdots, z_{n-1})$ is completed by joining the various pieces by vertical lines.

It is clear from the definitions that the lenth of $\Gamma(z_1, \cdots, z_{n-1})$ for $|z_n| \leqq b$ is $<$ const. b^2. Thus, for any $f \in \mathcal{D}$ which vanishes for $x_n > a$ we have

$$(5.35) \qquad \begin{aligned} \int \cdots \int dz_1 \cdots dz_{n-1} \int_{\Gamma(z_1, \cdots, z_{n-1})} [F(z)/J(z)]\, dz \\ = \int_{R'} [F(z)/J(z)]\, dz \\ = P \cdot f \end{aligned}$$

as in the case $n = 1$.

To prove that P is C^∞ in x_1 outside a neighborhood of 0, we have to show that we can find constants b_r so that for all real $z_1, \cdots, z_{n-1}$ and all $z_n \in \Gamma(z_1, \cdots, z_{n-1})$ we have

$$(5.36) \qquad |z_1{}^r \exp(-|\mathfrak{J}(z_n)|) \leqq b_r(1 + |z|).$$

Let us consider $z_1{}^r \exp(-|\mathfrak{J} z_n|)$ for $z_n \in \Gamma(z_1, \cdots, z_{n-1})$. Since J satisfies (1.1) and (1.2), we can find a $c_r > 0$ so large that for all $z' \in R$, if $|z'| > c_r$, then no point of the circle $|z''_n - z'_n| < 2r \log(|z'| + 1) + 1$

can contain a zero of $J_{z'_1,\cdots,z'_n}(z''_n)$ unless z' belongs to a set R''' on which $|z'_1|^r < d_r(1+|z|)$. In particular, this means that (for r large enough) the imaginary part of any point in $\Gamma(z_1,\cdots,z_{n-1})$ will be $> r\log|z'|$ except for $z'\in R'''$ or for $|z'|\leqq c_r$. This means that

$$(5.37) \qquad |z_1|^r \exp(-|\mathfrak{J}z|) \leqq d_r(1+|z|)+c_r^r$$

which implies inequality (5.36).

We can now proceed exactly as in the case $n=1$ to show that P is C^∞ in x_1 for $x_n > a$. Similar results hold for $x_n < -a$ and $x_j > a$ or $x_j < -a$, that is, P is C^∞ in x_1 for $|x| > a$.

Finally, P is a C^∞ in x_1 parametrix for S. For, we have for any $f\in D$,

$$\begin{aligned} S*P\cdot f = P\cdot S*f &= \int_{R'}[J(z)F(z)/J(z)]\,dz \\ &= \int_R F(z)\,dz - \int_{R''}F(z)\,dz \\ &= \delta\cdot f - \int_{R''}F(z)\,dz. \end{aligned}$$

If we set $W\cdot f = \int_{R''}F(z)\,dz$, then W is clearly a distribution. Moreover, by similar calculations as above, inequality (5.33) implies that W is C^∞ in x_1. This completes the proof of Proposition 5.7.

Putting the above Propositions 5.6, 5.7, 5.3 together we have (see Theorem III of the Introduction)

THEOREM 5.8. *Let $S\in\mathcal{E}'$ be invertible for $\mathcal{D}'$. Then* (1.1) *and* (1.2) *are a set of necessary and sufficient conditions for S to be C^∞ elliptic in x_1.*

A similar method applies to show

THEOREM 5.9. *Let $S\in\mathcal{E}'$ be invertible for $\mathcal{D}'_F$. Then S is weakly C^∞ elliptic in x_1 if and only if we can find an $m>0$ so that for each $r>0$ we can find a $b_r>0$ with the property that*

$$(5.38) \qquad |\mathfrak{J}z| \geqq m\log(1+|z_1|)$$

whenever $z\in V$ and

$$(5.39) \qquad |z_1^r| \geqq b_r(1+|z|).$$

THEOREM 5.10. *Let $S\in\mathcal{E}'$ be invertible for $\mathcal{D}'$. Then S is entire elliptic in x_1 if and only if we can find an $m>0$ so that*

$$(5.40) \qquad |\mathfrak{J}z| \geqq m(1+|z_1|)$$

whenever $z \in V$ *and*

$$|z_1| \geqq m^{-1} \log(1 + |z|). \tag{5.41}$$

We can also use similar methods to prove the analogs for relative ellipticity:

THEOREM 5.11. *Let* $S \in \mathcal{E}'$ *be invertible for* $\mathcal{D}'$. *Then* S *is* C^∞ *elliptic in* $(x_1, \cdots, x_l)$ *relative to* $(x_{l+1}, \cdots, x_n)$ *if and only if for each* $r \geqq 0$ *we can find a* $b_r > 0$ *with the property that*

$$|\mathfrak{J} z| \geqq r \log(1 + |(1 + |(z_1, \cdots, z_l)|) \tag{5.42}$$

whenever $z \in V$ *and*

$$|(z_1, \cdots, z_l)|^r \geqq b_r (1 + |(z_{l+1}, \cdots, z_n)|)^{b_r} (1 + |z|). \tag{5.43}$$

THEOREM 5.12. *Let* $S \in \mathcal{E}'$ *be invertible for* $\mathcal{D}'_F$. *Then* S *is weakly* C^∞ *elliptic in* $(x_1, \cdots, x_l)$ *relative to* $(x_{l+1}, \cdots, x_n)$ *if and only if we can find an* $m > 0$ *so that for each* $r > 0$ *we can find a* $b_r > 0$ *with the property that*

$$|\mathfrak{J} z| \geqq m \log(1 + |(z_1, \cdots, z_l)|) \tag{5.44}$$

whenever $z \in V$ *and*

$$|(z_1, \cdots, z_l)|^r \geqq b_r (1 + |(z_{l+1}, \cdots, z_n)|)^{b_r} (1 + |z|). \tag{5.45}$$

COROLLARY 5.12. *Let* $S \in \mathcal{E}'$ *be invertible for* $\mathcal{D}'$. *Suppose that in* x_1, S *is a differential operator with leading coefficient* 1. *Then* S *is weakly* C^∞ *elliptic in* x_1 *relative to* $(x_2, \cdots, x_n)$. *If* ∂ *is a partial differential operator in all variables and* $x_1 = 0$ *is non characteristic for* ∂, *then* ∂ *is* C^∞ *elliptic in* x_1 *relative to* $(x_2, \cdots, x_n)$.

Proof. Our hypotheses imply $J = z_1^p + \sum J_j z_1^j$, where $J_j \in \boldsymbol{E}'(z_2, \cdots, z_n)$. Then for fixed $(z_2, \cdots, z_n)$, $J(z)$ does not vanish if

$$|z_1| \geqq \text{const.} \max |J_j(z_2, \cdots, z_n)|.$$

Now, for some $A > 0$

$$\max |J_j(z_2, \cdots, z_n)| \leqq A(1 + |(z_2, \cdots, z_n)|)^A \exp(A\, |\mathfrak{J}(z_2, \cdots, z_n)|).$$

Thus, $J(z)$ does not vanish if

$$|z_1| \geqq \text{const}(1 + |z_2, \cdots, z_n|)^A \exp(A\, |\mathfrak{J}(z_2, \cdots, z_n)|).$$

By Theorem 5.12 this implies that J is weakly C^∞ ellipitic in x_1 relative to $(x_2, \cdots, x_n)$.

The proof for θ is similar.

THEOREM 5.13. *Let S in $\mathcal{E}'$ be invertible for $\mathcal{D}'$. Then S is entire elliptic in $(x_1,\cdots,x_l)$ relative to $(x_{l+1},\cdots,x_n)$ if and only if we can find an $m>0$ so that*

$$|\mathcal{I}z|\geqq m(1+|(z_1,\cdots,z_l)|) \tag{5.46}$$

whenever $z\in V$ and

$$|(z_1,\cdots,z_l)|\geqq m^{-1}\log(1+|z|). \tag{5.47}$$

That is, relative entire ellipticity is the same as entire ellipticity.

Remark. We could have defined relative ellpiticity in terms of convolution by elements of Carleman non quasi-analytic classes instead of $\mathcal{D}$. In this case, the right hand side of (5.47) would be changed and the class of relative entire elliptic operators would presumably be different for some Carleman classes, though I have not constructed any examples.

In case S is invertible, we can show, using the methods of Section 2, that every distribution W which is C^∞ (entire) in x_1 can be written in the form $S*T$, where T is again C^∞ (entire) in x_1. Thus, if S is invertible, then every solution T of the equation $S*T=W$, where W is C^∞ (entire) in x_1, is also C^∞ (entire) in x_1 provided we know that every solution T of $S*T=0$ is C^∞ (entire) in x_1. Thus, if S is invertible, then the conditions for ellipticity can be stated in terms of the homogeneous equation $S*T=0$.

I do not know if there exists an S which is not invertible and is C^∞ elliptic in x_1. However, in case we are considering C^∞ (entire) ellipticity in all variables, we shall see that no S with the corresponding property can exist without being invertible, that is, if for all $f\in\mathcal{E}$ ($f\in\mathcal{H}$) all the distribution solutions T of

$$S*T=f$$

are again in $\mathcal{E}$ (in $\mathcal{H}$), then S is invertible.

Definition. J is called *extra slowly decreasing in* z_1 if for each $a>0$ and each $k>0$ there exists an $m>0$ so large that

$$\liminf_{|\mathcal{I}(z)|\leqq a+a\log(1+|\mathcal{R}z|)}|z_1|^m|J(z)|=\infty. \tag{5.48}$$

THEOREM 5.14. *If J is not extra slowly decreasing in z_1, then there exists a $T\in\mathcal{D}'$ so that T is not C^∞ in x_1 but $S*T$ is C^∞ in x_1. In particular, if S is C^∞ elliptic in all variables, then S is invertible.*

Proof. Assume J is not extra slowly decreasing in z_1. Then we can find a $k>0$ and a sequence $\{_jz\}$ with $|\mathfrak{J}(_jz)|\leqq a+a\log(1+|\mathfrak{R}(_jz)|)$, $|_jz|\to\infty$, $|_jz_1|^k>|_jz|$ and for each m,

$$\limsup |_jz_1|^m|J(_jz)|=M_m<\infty. \tag{5.49}$$

As in the proof of Proposition 5.3, the series $\sum\delta_{j^z}$ converges in $\mathfrak{D}'$ to W whose Fourier transform T is not C^∞ in x_1. But using

$$JW=\sum J(_jz)\delta_{j^z} \tag{5.50}$$

we see easily that $\{z_1{}^m(M_m)^{-1}JW\}$ is bounded on the bounded sets of $\mathfrak{D}$, hence, is bounded in $\mathfrak{D}'$. Thus, $S*T$ is C^∞ in x_1 which proves our assertion.

Remark 1. We could easily make T to be a q times differentiable function on the set $|x|\leqq q$.

Remark 2. The above theorem, as well as the succeeding ones, can be easily extended to the cases where "distribution" is replaced by "distribution of finite order," and "C^∞ in x_1" is replaced by "entire in x_1."

From Theorems 5.14 and 5.8 we deduce immediately

THEOREM 5.15. *A necessary and sufficient condition that S be C^∞ elliptic in all variables is:*

$$\liminf_{z\in V,|z|\to\infty}|\mathfrak{J}z|/\log(1+|z|)=\infty. \tag{5.51}$$

Remark. It seems that the conditions: $S*T=0$ implies T is C^∞ in x_1, and J is extra slowly decreasing in z_1 should imply that S is C^∞ elliptic in x_1. However, I do not know how to decide this.

I should now like to give several examples:

Example 1. We give an example of an S which is entire elliptic in x_1 but is not C^∞ elliptic in all variables. A trivial example is $S=\partial/\partial x_1$. A less trivial examples is $S=\partial/\partial_1-i\partial/\partial x_2-i\partial/\partial x_3$. Then $J(z)=iz_1+iz_2+z_3$. Write $z_j=\xi_j+i\eta_j$. Then $J(z)=0$ is equivalent to

$$\xi_1=-\eta_2-\eta_3$$

$$\eta_1=\xi_2+\xi_3.$$

Now, $|\mathfrak{J}z|=|\eta_1|+|\eta_2|+|\eta_3|$, so that

$$|\mathfrak{J}z/z_1|=|\eta_1/z_1|+(|\eta_2|+|\eta_3|)/|z_1|. \tag{5.32}$$

If $|\eta_1|\leqq\frac{1}{2}|z_1|$, then $|\xi_1|>\frac{1}{2}|z_1|$. Thus

$$(5.53)\qquad (|\eta_2|+|\eta_3|)/|z_1| \geqq |\eta_2+\eta_3|/|z_1| = |\xi_1|/|z_1| > \tfrac{1}{2}.$$

This combined with (5.52) shows that for all $z \in V$, $|\mathfrak{J}z|/|z_1| \geqq \frac{1}{2}$. Hence, by Theorem 5.10 S is entire elliptic in x_1.

Since $V \cap R$ is not compact, S is clearly not C^∞ elliptic in all variables.

Example 2. I want to construct an example of an $S \in \mathfrak{D}$ which satisfies (1.1) and (1.2). Since $S \in \mathfrak{D}$, S cannot be invertible and S cannot be C^∞ ellipitic.

Let us consider first the case $n=1$. We can construct an $F \in \mathfrak{D}$ which is even, $F(0)=1$, $F(x)=O(\exp(-|x|^{\frac{1}{2}})$; the possibility of constructing such an F is well-known from the theory of quasi-analytic functions. We write

$$(5.54)\qquad F(z)=\pi(1-z^2/a_j^2).$$

We may assume, by replacing a_j by $|a_j|$ if necessary, that the a_j are real and positive, because if we replace a_j by $|a_j|$, then for any real x, we have

$$\begin{aligned} |1-x^2/|a_j|^2| &= |a_j|^{-2}\,|\,|a_j|^2-x^2| \\ &\leqq |a_j|^{-2}\,|a_j^2-x^2| \\ &= |1-x^2/a_j^2|. \end{aligned}$$

Thus, the infinite product (5.54) does not increase for $z \in R$ when we replace a_j by $|a_j|$.

Next we define

$$(5.55)\qquad J(z)=\pi(1-z^2/(a_j^2+ia_j^{\frac{5}{4}}).$$

For j large, the imaginary part of $(a_j^2+ia_j^{\frac{5}{4}})^{\frac{1}{2}}$ is about $a_j^{\frac{1}{4}}$. Thus it remains to show that $J \in \boldsymbol{D}$.

Actually, it is very difficult to show that $J \in \boldsymbol{D}$ by comparing it with F directly. However, it follows easily from the minimum modulus theorem that

$$(5.56)\qquad F((x^2+i|x|^{\frac{5}{4}})^{\frac{1}{2}})=O(\exp(-\tfrac{1}{2}|x|^{\frac{1}{2}}).$$

(Since F is even, it does not matter which square root we take on the left.) There is much more hope in showing $J(x)/F((x^2+i|x|^{\frac{5}{4}})^{\frac{1}{2}})$ is bounded and so to conclude that $J \in \boldsymbol{D}$.

We have for $x>0$,

$$(5.57)\qquad \begin{aligned} J(x)/F((x^2+ix^{\frac{5}{4}})^{\frac{1}{2}}) &= \prod(1-x^2/(a_j^2+ia_j^{\frac{5}{4}}))/\prod(1-(x^2+ix^{\frac{5}{4}}/a_j^2)) \\ &= \prod((a_j^2+a_j^{\frac{5}{4}}-x^2)/(a_j^2-(x^2+ix^{\frac{5}{4}}))) \cdot \prod(a_j^2/(a_j^2+ia_j^{\frac{5}{4}})). \end{aligned}$$

The reciprocal of the last product is

$$\prod((a_j^2+ia_j^{\frac{5}{4}})/a_j^2)=\prod(1+ia_j^{-\frac{3}{4}}).$$

It is well known (see e.g. [29]) that this latter product converges absolutely. Hence, so does $\prod(a_j^2/(a_j^2+ia_j^{\frac{5}{4}}))$.

We are thus left to consider the product

$$\prod(a_j^2+ia_j^{\frac{5}{4}}-x^2)/(a_j^2-x^2-ix^{\frac{3}{2}}).$$

For $x^{\frac{3}{2}}\geqq a_j^{\frac{5}{4}}$, the terms of the product are all of modulus $\leqq 1$. If $x^{\frac{3}{2}}<a_j^{\frac{5}{4}}$ and if $a_j>a_0$ (independent of x), then

$$\begin{aligned}|(a_j^2+ia_j^{\frac{5}{4}}-x^2)/(a_j^2-x^2-ix^{\frac{3}{2}})|^2 &= ((a_j^2-x^2)^2+a_j^{\frac{5}{2}})/((a_j^2-x^2)^2+x^3)\\ &(a_j^2-x^2)^2/((a_j^2-x^2)^2+x^3)+a_j^{\frac{5}{2}}/((a_j^2-x^2)^2+x^3)\\ &\leqq 1+2a_j^{-\frac{3}{2}}.\end{aligned}$$

Hence, except possibly for a polynomial factor in x,

$$\prod|(a_j^2+ia_j^{\frac{5}{4}}-x^2)/(a_j^2-x^2-ix^{\frac{3}{2}})|\leqq M,$$

where M is independent of x. This proves that $J\in\boldsymbol{D}$, which is the desired result.

In case $n>1$, we define J_1 as follows:

$$J_1(z)=J((z_1^2+z_2^2+\cdots+z_n^2)^{\frac{1}{2}}), \tag{5.58}$$

where J is defined as in (5.55). Note that since J is even, J_1 is entire. It is immediately verified that $J_1\in\boldsymbol{D}$. We want to examine the zeros of J_1. If z is such a zero, then for some j,

$$z_1^2+z_2^2+\cdots+z_n^2=a_j^2+ia_j. \tag{5.59}$$

We write $z_k=\xi_k+\eta_k$. Then (5.59) becomes

$$\begin{aligned}\xi_1^2+\xi_2^2+\cdots+\xi_n^2-\eta_1^2-\eta_2^2-\cdots-\eta_n^2&=a_j^2\\ 2\xi_1\eta_1+2\xi_2\eta_2+\cdots+2\eta_n\xi_n&=a_j^{\frac{5}{4}}\end{aligned} \tag{5.60}$$

or

$$\begin{aligned}|\xi|^2-|\eta|^2&=a_j^2\\ 2|\xi\cdot\eta|&=a_j^{\frac{5}{4}}\end{aligned} \tag{5.61}$$

Now, if $|\xi|>2|\eta|$, then (5.61) shows that $|\xi|^2-\frac{1}{4}|\xi|^2\leqq a_j^2$, so that $|\xi|\leqq 2a_j$. Hence, by (5.61) again,

$$a_j^{\frac{5}{4}}\leqq 2|\xi|\cdot|\eta|\leqq 4a_j|\eta|.$$

Thus, $|\eta|\geqq\frac{1}{4}a_j^{\frac{1}{4}}$, which is the desired result.

Remark. I suspect that all distribution solutions of $S*T=0$ are in $\mathcal{E}$, but I cannot prove this. In case $n=1$, it may be possible to prove this by use of Schwartz' mean-periodic expansion.

Example 3. We give an example of an $S\in\mathcal{E}'$ such that all distribution solutions of $S*W=0$ are in C^∞ but S is not invertible, hence, S is not C^∞ elliptic: Let $n=1$, and let S_1 be C^∞ elliptic and let J_1 have infinitely many zeros; we may assume J is even. Choose a very lacunary infinite sequence of the zeros a_j of J_1, so lacunary that $J_2(z)=\prod(1-z^2/a_j^2)$ is of zero order, or even that if $M_2(r)=\max\limits_{|z|=r}|J_2(z)|$, then $\log M_2(r)=O((\log r)^2)$. Then (see Boas, *Entire Functions*, p. 50, the proof of Theorem 3.6.1), if $m_2(r)$ denotes the minimum of $J_2(z)$ on a circle of radius r, we will have $\log m_2(r) \geqq \frac{1}{2}\log M_2(r)$ except on a sequence of intervals $I_j=\{b_j\leqq r\leqq c_j\}$, where each I_j contains some $|a_j|$ and where $\sum\limits_{c_j\leqq R}(c_j-b_j)\leqq\frac{1}{2}R$ for R sufficiently large.

Suppose we choose the a_j in such a way that $|a_{j+1}|>(10+|a_j|)^2$ for all j. For j large enough, if we call $x_j=|a_j|/3$, then for any $y\in R$ with $|y-x_j|<|x_j|/6$ we have

$$|J_2(y)|\geqq m_2(|y|)\geqq\tfrac{1}{2}M_2(|y|).$$

On the other hand, by Liouville's theorem, for all p, $\liminf M_2(r)/r^p=\infty$. It follows that $J(z)=J_1(z)/J_2(z)$ cannot be slowly decreasing. This can be seen fairly easy by applying the minimum modulus as above to the functions $J_2(z)/(1-z^2/a_j^2)$ from which it is deduced that except for the effect of the factors $(1-z^2/a_j^2)$ near a_j, the division of J_1 by J_2 serves to decrease $|J_1|$. The term $1-z^2/a_j^2$ is handled by using the fact that for $|z|\geqq 4$ we have $|1-z^2/a_j^2|\geqq|z|^{-2}$ except in circles of radius 4 about $\pm a_j$; these circles can be treated by the maximum modulus theorem. This argument shows that, in fact, $J\in\boldsymbol{D}$.

Next we note that J_2 is not in $\boldsymbol{E}'$. However, using the methods of [14] we can find a space B of functions of compact support such that the inverse Fourier transform S_2 of J_2 is well defined on B.

Now, suppose $T\in\mathcal{D}_1$ satisfies $S*T=0$. We consider T as an element of B' and it must therefore satisfy, as an element of B',

$$0=S_2*S*T=S_1*T.$$

Since B is dense in $\mathcal{D}$ this implies $S_1*T=0$ as an element of $\mathcal{D}'$. Hence, $T\in\mathcal{E}$ because S_1 is C^∞ elliptic. This proves our assertion.

For $n>1$ we could proceed as in example 2 above.

Example 4. I give an example of an $S \in \mathcal{E}'$ which satisfies (1.1) and (1.2) in all variables but for which there exists a distribution T satisfying $S * T = 0$, with $T \notin \mathcal{E}$. We assume $n = 1$; the passage to $n > 1$ is as in Example 2 above.

Let us define $J_1(z) = \prod \cos(z/j!\,j\log^2 j)^{j!}$. Let us choose $J_2 \in \boldsymbol{D}$ so that $J_2(x)\exp(|x|^{2/3})$ is bounded on R. (The existence of J_2 is guaranteed by the Denjoy-Carlemann theorem on quasi-analytic classes.) Let $J_3(z) = J_1(z)J_2(z)$. As in Example 2 above, we can "shift" the zeros of J_3 to the curve $\mathfrak{J}z = |Rz|^{\frac{1}{2}}$ and obtain a function $J \in \boldsymbol{D}$.

J has a zero at $j!\,j\log^2 j + i(j!\,j)^{\frac{1}{2}}\log j$ of order $j!$. Thus, if s denotes the Fourier transform of J, then we have solutions of the equation $S * f = 0$ of the form

$$f_j(x) = Q_j(x)\exp(ixj!\,j\log^2 j - x(j!\,j)^{\frac{1}{2}}(\log j)^{\frac{1}{2}}),$$

where Q_j is any polynomial of degree $\leqq j!$. Of course, I want to choose Q_j in a suitable manner. If all the solutions of $S * T = 0$ were in E, then we could find a $b > 0$ so large that

$$f'_j(0) \leqq b \max_{|x| \leqq b} f_j(x). \tag{5.62}$$

This is a consequence of the closed graph theorem which shows that δ' (the derivative of the δ) is continuous on the subspace of $\mathcal{E}$ of solutions of S in the topology induced by the space of continuous functions on R.

Thus, I want to choose Q_j to violate (5.62). For this purpose I am led to the following problem: Let P_k be a polynomial of degree k such that $|P_k(x)|\exp g_k|x| \leqq 1$ for $|x| \leqq 1$. Suppose $p_k(0) = 0$ and $|P'_k(0)| \geqq 1$; then how large can g_k be? We assert that we can take $g_k \geqq \text{const}\, k^{\frac{1}{2}}$.

Before proving this, I show how it can be used to prove our result: Inequality (5.62) shows that we can find a $b > 0$ so that for any polynomial Q_j of degree $\leqq j!$ with $Q_j(0) = 0$ we have

$$\begin{aligned} &|Q'_j(0)|\,|ij!\,j\log^2 j - (j!\,j)^{\frac{1}{2}}(\log j)^{\frac{1}{2}}| \\ &\qquad \leqq b \max_{|x| \leqq b} |Q_j(x)|\exp((j!\,j)^{\frac{1}{2}}(\log j)^{\frac{1}{2}}|x|). \end{aligned} \tag{5.63}$$

I claim we cannot have $b = 1$ and the general case follows by the simple transformation $x \to x/b$.

Our above construction of the P_k shows we can choose Q_j so that

$$\max_{|x| \leqq 1} |Q_j(x)|\exp((j!)^{\frac{1}{2}}|x|) \leqq 1$$

but $|Q'_j(0)| \geqq 1$. That is, the right side of (5.63) is $\leqq 1$ but the left side is arbitrarily large. This is a contradiction and proves our result.

Remark. We could actually construct a $T \in \mathcal{D}'$ which satisfies $S * T = 0$, but $T \notin \mathcal{E}$; we could even make T differentiable as often as we want.

It remains to prove our assertion on the existence of P_k: We pick $P_{2k+1} = x(1-x^2)^k$. Then clearly $P_k(0) = 0$, $P'_k(0) = 1$. If we could show that $\max\limits_{0 \leqq x \leqq 1} (1-x^2)^k e^{\frac{1}{2}x}$ is bounded from above uniformly in k, then clearly so is $\max\limits_{|x| \leqq 1} | P_{2k+1}(x) | e^{k^{\frac{1}{2}}|x|}$.

The derivative of $(1-x^2)^k e^{k^{\frac{1}{2}}x}$ is

$$k^{\frac{1}{2}}(1-k^2)^k e^{k^{\frac{1}{2}}x} - 2kx(1-x^2)^{k-1} e^{k^{\frac{1}{2}}x}.$$

This vanishes (for $k > 1$) if $x = \pm 1$ or if

$$1 - x^2 - 2k^{\frac{1}{2}}x = 0,$$

$$x^2 + 2k^{\frac{1}{2}}x - 1 = 0$$

$$x = -k^{\frac{1}{2}} \pm (k+1)^{\frac{1}{2}}.$$

The value we are interested is thus $x = -k^{\frac{1}{2}} + (k+1)^{\frac{1}{2}}$. Since $(1-x^2)^k e^{k^{\frac{1}{2}}x}$ vanishes at $x = 0, 1$, is non-negative in the interval $[0,1]$ and has $x = -k^{\frac{1}{2}} + (k+1)^{\frac{1}{2}}$ as its only critical point in this interval, it follows that $x = -k^{\frac{1}{2}} + (k+1)^{\frac{1}{2}}$ is a maximum. The value of the function at this point is

$$(1 - ((k+1)^{\frac{1}{2}} - k^{\frac{1}{2}})^2)^k e^{k^{\frac{1}{2}}((k+1)^{\frac{1}{2}} - (k^{\frac{1}{2}}))}.$$

Note that $(k+1)^{\frac{1}{2}} - k^{\frac{1}{2}} < \frac{1}{2}$ for, squaring, we have to show that

$$k + 1 < \tfrac{1}{4} + k + k^{\frac{1}{2}}$$

which is clear. Moreover, $(k+1)^{\frac{1}{2}} - k^{\frac{1}{2}} > \frac{1}{4}k^{-\frac{1}{2}}$ for squaring we must show that $k + 1 > k + \frac{1}{16}k^{-1} + \frac{1}{2}$ which is again clear. Thus,

$$(1 - ((k+1)^{\frac{1}{2}} - k^{\frac{1}{2}})^2)^k e^{k^{\frac{1}{2}}((k+1)^{\frac{1}{2}} - k^{\frac{1}{2}})} \leqq (1 - \tfrac{1}{16}k^{-1})^k e^{\frac{1}{2}k^{\frac{1}{2}}}. \tag{5.64}$$

The right side of (5.64) behaves for large k like $\exp(-\frac{1}{16}k + \frac{1}{2}k^{\frac{1}{2}}) \to 0$. Thus the left side of (5.64) is uniformly bounded in k which completes the proof of our assertion.

Remark. It would be of interest to find the best possible g_k and also the corresponding P_k.

I wish now to show that those S which are entire elliptic in all variables are just (essentially) the classical elliptic differential operators. More generally, we have

THEOREM 5.16. *Let S be entire elliptic in x_1. Then, in x_1, S is the composition of a differential operator with a translation, that is, we can find a real number a and a finite sequence $\{S_j\}$ of distributions which are independent of x_1 so that*

$$S = \sum S_j \times (\partial^j/\partial x_1{}^j * \tau_a). \tag{5.65}$$

Here $\times$ denotes the direct product of distributions and τ_a is translation by a in the x_1 direction.

Proof. Let $b_2, b_3, \cdots, b_n$ be fixed complex numbers such that $J(z_1, b_2, \cdots, b_n)$ does not vanish identically. By Proposition 5.3. the zeros of $J(z_1, b_2, \cdots, b_n)$ lie (except for a finite number) outside of an angular segment containing the real axis. Carleman's theorem (see [29]) this means that the density of zeros of $J(z_1, b_2, \cdots, b_n)$ is zero. Thus, $J(z_1, b_2, \cdots, b_n)$ is a polynomial in z_1 times a pure imaginary exponential in z_1, that is, for some a

$$J(z_1, b_2, \cdots, b_n) = \sum_{j=1}^{r} J_j(b_2, \cdots, b_n) z_1{}^j \exp(iaz_1), \tag{5.66}$$

where $J_j(b_2, \cdots, b_n)$ are certain complex numbers.

A priori, both a and r might depend on $b_2, \cdots, b_n$. If we multiply $J(z_1, b_2, \cdots, b_n)$ by $J(-z_1, b_2, \cdots, b_n)$ we obtain a new function $J^1(z_1, b_2, \cdots, b_n)$ of the form

$$J^1(z_1, b_2, \cdots, b_n) = \sum_{j=1}^{2r} J_j{}^1(b_2, \cdots, b_n) z_1{}^j. \tag{5.67}$$

Here again, r may depend on $b_2, \cdots, b_n$. But, if we apply Baire's category theorem, we can find an open set of $(b_2, \cdots, b_n)$ on which r is bounded, say by r_0. We expand both sides of (5.64) in power series in z_1 and it follows by analyticity that any $J_j{}^1$ must vanish if $j > 2r_0$. Thus, r is bounded.

Next, all the $J_j{}^1$ are entire functions of exponential type. It follows easily by comparing coefficients that the J_j are also entire functions in $\boldsymbol{E}'$. Hence, since a is bounded by the exponential type of J, a is an entire function of $(b_2, \cdots, b_n)$ and so must be a constant.

Using the result that $J^1(z_1, b_2, \cdots, b_n)$ is a polynomial of degree $\leqq 2r_0$ in z_1 we could conclude that a is a constant by applying the theorem of addition of supports (see [24]). It then again follows immediately that the J_j are entire functions of exponential type which lie in $\boldsymbol{E}'$. This completes the proof of Theorem 5.16.

Next, we want to find those differential difference operators which are C^∞ elliptic in x_1. We shall see that they are essentially differential operators in x_1. More generally, we have

THEOREM 5.17. *Let S be a differential difference operator in x_1 which is C^∞ elliptic in x_1. Then, in x_1, S is the composition of a differential operator with a translation.*

Proof. As in the proof of Theorem 5.16, we fix complex numbers $b_2, \cdots, b_n$ such that $J(z_1, b_2, \cdots, b_n)$ is identically zero. We could conclude the proof the same way as in the above theorem if we could prove that $J(z_1, b_2, \cdots, b_n)$ is a polynomial in z_1 times a pure imaginary exponential in z_1. This is a consequence of Theorem 5.8 and

LEMMA 5.18. *Let Q be an exponential polynomial in one variable with pure imaginary exponentials such that, if ${}_jz$ denotes its zeros, then*

$$\liminf |\mathfrak{I}({}_jz)|/\log(1 + |{}_jz|) = \infty.$$

Then Q is a polynomial times a pure imaginary exponential.

Proof. We write $Q(z) = \sum_{k=1}^{s} P_k(z)\exp(ib_kz)$, where P_k are polynomials not identically zero and b_k are real numbers with $b_1 < b_2 < \cdots < b_s$. If Q has a finite number of zeros the result is easy. If not, there exists a sequence of complex numbers c_l, with $|c_l| \to \infty$, which are zeros of Q. We show that this is impossible.

We may suppose for simplicity that $\mathfrak{I}c_l > 0$ for all l. Then we have

$$\begin{aligned} |P_s(c_l)| &= |\sum_{k<s} P_k(c_l)\exp(ic_l(b_k - b_s))| \\ &\leq c(1 + |c_l|^m)\exp(-(\mathfrak{I}c_l)(b_{s-1} - b_s)) \end{aligned} \tag{5.68}$$

for suitable c, m which are independent of l. By (5.68) it follows that $P_s(c_l) \to 0$ which is impossible since P_s is a polynomial not identically zero.

This completes the proof of Lemma 5.18 and hence of Theorem 5.17.

Remark. The result corresponding to Theorem 5.17 for weak C^∞ ellipticity does not hold as the example $(n = 1)S = d/dx - \tau$ shows. For then $J = iz - \exp(iz)$. For $J(z) = 0$ we have $(z = \xi + i\eta)$

$$|iz| = \exp(-\eta)$$

so

$$\xi^2 + \eta^2 = \exp(-2\eta). \tag{5.69}$$

Hence, if $\eta \leqq 0$, we have

$$-2\eta \geqq 2 \log |\xi|$$

so

$$|\eta| \geqq \log |\xi|.$$

If $\eta > 0$, then inequality (5.69) defines a compact set in the z plane so there are only a finite number of zeros of J there. Thus, S is weakly C^∞ elliptic. The fact that S is weakly C^∞ elliptic can also be seen easily directly. On the other hand, $S * \sum \delta_j^{(j)} = 0$ where $\delta_j^{(j)}$ is the j-th derivative of the unit mass at the point 1; this shows again that S is not C^∞ elliptic.

For $n > 1$, a similar computation shows that if

$$J = \exp(i(z_1 + z_2 + \cdots + z_n)) - (z_1^2 + z_2^2 + \cdots + z_n^2)$$

then S, which is a differential operator, is weakly C^∞ elliptic in all variables.

6. Unsolved problems and general remarks. In addition to the problems and remarks stated in the text, we have the following:

1. One of the most important problems is to give, in case $n = 1$, conditions on the zeros of a $J \in \boldsymbol{E}'$ to insure that J should be slowly decreasing. Certain sufficient conditions are known if the zeros are "close to" the integers (see, e. g., [22]). But all these results are obtained by reducing the question to the known case of $\sin z$. Certainly a necessary and sufficient condition would be of great interest. In this connection we have one partial result:

PROPOSITION 6.1. *Suppose J has a sequence of zeros a_j of multiplicities r_j which satisfy*

$$\liminf r_j / (|\mathfrak{I} a_j| + \log |\mathfrak{R} a_j|) = \infty. \tag{6.1}$$

Then J is not slowly decreasing.

Proof. We may suppose as usual that $|J(z)| \leqq 1$ for $z \in R$; suppose for simplicity that J is of exponential type $\leqq 1$. Then the Phragmén-Lindelöf theorem (see [29] tells us that on the line $\mathfrak{I} z = \mathfrak{I} a_j$ we have $|J(z)| \leqq \exp(|\mathfrak{I} a_j|)$. We now apply Bernstein's theorem (see, e.g., [14]) which shows that

$$|J^{(k)}(a_j)| \leqq \exp(|\mathfrak{I} a_j|). \tag{6.2}$$

Now, let us examine the Taylor expansion of J about a_j:

$$
\begin{aligned}
|J(z)| &= |\sum J^{(k)}(a_j)(z-a_j)^k/k!| \\
&\leqq \exp(|\mathfrak{I}a_j|)\sum_{k\geqq r_j}|z-a_j|^k/k! \\
(6.3)\qquad &= \exp(|\mathfrak{I}a_j|)|z-a_j|^{r_j}\sum_{k\geqq r_j}|z-a_j|^{k-r_j}/k! \\
&\leqq (r_j!)^{-1}\exp(|\mathfrak{I}a_j|)|z-a_j|^{r_j}\sum_{k\geqq 0}|z-a_j|^k/k! \\
&\leqq (r_j!)^{-1}\exp(|\mathfrak{I}a_j|)|z-a_j|^{r_j}\exp|z-a_j|.
\end{aligned}
$$

Thus, $|J(z)|\leqq|\mathfrak{R}a_j|^{-l}$ whenever

$$|z-a_j|^{r_j}\leqq|\mathfrak{R}a_j|^{-l}\exp(-|\mathfrak{I}a_j|)-|z-a_j|)(r_j)!,$$

that is, whenever

$$r_j\log|z-a_j|\leqq -l\log|\mathfrak{R}a_j|-|\mathfrak{I}a_j|-|z-a_j|+\log r_j!.$$

Using Stirling's formula, this is true whenever

$$r_j\log|z-a_j|+|z-a_j|-r_j\log r_j\leqq -l\log|\mathfrak{R}a_j|-|\mathfrak{I}a_j|-\text{const.}\,r_j.$$

We now use our hypothesis (6.1) and we obtain the result that $|J(z)|\leqq|\mathfrak{R}a_j|^{-l}$ whenever j is large and

$$(6.4)\qquad r_j\log|z-a_j|+|z-a_j|-r_j\log r_j\leqq -\text{const.}\,r_j.$$

It is clear that (6.4) is satisfied whenever $|z-a_j|\leqq br_j$ for a suitable $b>0$. This shows that J cannot be slowly decreasing, which is our assertion.

It is clear that inequality (6.1) can be slightly ameliorated, and moreover, that we don't need r_j zeros at a_j, but only sufficiently close to a_j (in that case the first r_j Taylor coefficients of J at a_j are very small and the others are handled as before). However, even an "almost sufficient" condition of this type for J to be slowly decreasing seems very difficult.

2. Presumably, we could use the results of Schwartz [25] to show that if J has real distinct zeros, is slowly decreasing, and if the index of condensation of the zeros is zero, then the quotient space $\boldsymbol{E}'/J\boldsymbol{E}'$ could be described as the space of slowly increasing functions (i.e. sequences) on the zeros of J. This method could also be slightly modified in case the zeros of J are not real and distinct. The problem arises as to what is the quotient space $\boldsymbol{E}'/J\boldsymbol{E}'$ in general. This problem is connected with my fundamental principle (see [15], [16]) and has been solved for some slowly decreasing J even in case $n>1$. But I know of no answer to the question in case J is not slowly decreasing even if $n=1$.

3. In connection with the results of Section 2, can Theorems 2.8, 2.9 be completed to describe completely $S * \mathfrak{D}$ and to decide when $S * \mathfrak{D}' \supset T * \mathfrak{D}'$? In particular, is $S * \mathfrak{D}$ bornologic?

4. Is $\mathfrak{D} * \mathfrak{D} = \mathfrak{D}$? Is $\mathfrak{D} * \mathcal{E} = \mathcal{E}$, or even $\mathfrak{D} * \mathfrak{D}' = \mathcal{E}$?

5. In Section 4 we proved that $\bigcap_{S \in \mathcal{E}'} S * \mathfrak{D}' = \bigcap_{S \in \mathcal{E}'} S * \mathcal{E} = A$. We can improve this result slightly as follows: Let M be any monotonic increasing sequence for which A_M is not quasi analytic. Define $\mathcal{E}_M$ as the subspace of $f \in \mathcal{E}$ which, together with all their derivatives, satisfy inequality (4.1) on every compact set. We could use the methods of Section 4 to show that $\bigcap_{S \in \mathcal{E}'} S * \mathcal{E}_M = A$. But, is $\bigcap_{S \in \mathcal{E}'} S * A = A$?

6. Closely related to problem 5 is the following problem: Let $\mathfrak{D}_M = \mathfrak{D} \cap \mathcal{E}_M$. We introduce a natural topology in $\mathfrak{D}_M$ as in [14]; call $\mathfrak{D}_M$ the dual of $\mathfrak{D}_M$. Given $S \in \mathcal{E}'$, can we always find an M such that the equation $S * T = \delta$ has a solution $T \in \mathfrak{D}'_M$, or even, can we find an M such that $S * \mathfrak{D}'_M \supset \mathfrak{D}'$?

We give now an example to show that this is not the case. Let $n = 1$ and define the Fourier transform J of S to be

$$J(z) = \prod \cos(z/j!\,j \log^2 j)^{j!}. \tag{6.5}$$

THEOREM 6.2. *S is not invertible for any $\mathfrak{D}'_M$ for which $\mathcal{E}_M$ is Carleman non quasi-analytic. In fact, for no such M does there exist a $T \in \mathfrak{D}'_M$ which satisfies $S * T = \delta$.*

Proof. If there were such a T, then for B any set in $\mathfrak{D}$ such that $S * B$ is bounded in $\mathcal{E}'$, we must have B bounded in $\mathcal{E}'_M$. For,

$$B = (S * T) * B = T * (S * B)$$

and so is bounded in $\mathcal{E}'_M$ by the continuity of convolution.

Let us set $n_j = j!\,j \log^2 j$. We return to the notation of the proof of Theorem 2.1. We set

$$L_j(z) = \exp(n_j/j \log j) H_{[n_j/j \log j]}(z - n_j) \tag{6.6}$$

and call $B = \{L_j\}$. Then I claim that JB is bounded in $\boldsymbol{E}'$ but B is not bounded in $\boldsymbol{E}'_M$ for any M such that $\mathcal{E}_M$ is non quasi-analytic.

The fact that JB is bounded in $\boldsymbol{E}'$ is seen as follows: JL_j is certainly small far away from n_j, for $|J(x)| \leqq 1$ for $x \in R$ and $|L_j(x)| \leqq R$ when $|x - n_j| \geqq n_j/j \log j$.

Now, $\cos(z/n_j)$ vanishes when $z = n_j$. Moreover, it behaves linearly near n_j with slope $1/n_j$. Thus we ask when is $(x/n_j)^{j!} = \exp(-n_j/j \log j)$, that is when is

$$j!(\log x - \log n_j) = -n_j/j \log j$$

$$j!(\log x - \log j - 2 \log\log j - \log j!) = -j! \log j$$

$$\log x = \log j + 2 \log\log j + \log j! - \log j$$

$$\begin{aligned} x &= j! \log^2 j \\ &= j!\, j \log^2 j/j \\ &= n_j/j. \end{aligned}$$

It follows easily that JB is bounded in $\boldsymbol{E}'$.

Next we show that B is not bounded in $\boldsymbol{E}'_M$ for $\mathcal{E}_M$ non quasi-analytic. Now, the bounded sets of $\boldsymbol{E}'_M$ can be described as follows: All the functions are of fixed exponential type and are majorized on R by a continuous monotonic increasing function $H \geqq 1$ for which

$$\int [\log H(x)/(1+x^2)]\,dx < \infty.$$

In particular, if B were bounded,

$$\log H(n_j) \geqq n_j/j \log j.$$

But the n_j are lacunary enough so that

$$\begin{aligned} \int_{n_j}^{n_j+1} [\log H(n_j)/(1+x^2)]\,dx &\geqq \log H(n_j)/2n_j \\ &\geqq 1/2j \log j. \end{aligned}$$

Thus,

$$\int [\log H(x)/(1+x^2)]\,dx \geqq \tfrac{1}{2} \sum 1/j \log j = \infty.$$

This contradiction completes the proof of Theorem 6.2.

7. In Section 5 we used lacunary series of exponentials to construct examples. It should be of interest to study these series in more detail in case $n > 1$.

8. The results of Section 2 are non constructive. In fact, it would be of interest to give a constructive method for finding an elementary solution for S in case S is invertible. It is not difficult to give such a procedure in

case S is a partial differential difference operator. Results of this kind are of importance in studying equations depending on a parameter (see, e.g., [28]).

9. Finally, the problem arises as to extend the results of this paper to systems of convolution equations. In case the determinant of the system $\neq 0$, then we can use the methods of this paper together with Cramer's rule to obtain the corresponding results. But, in case the determinant is $\equiv 0$, the problem seems to be extremely difficult. The simplest example is probably the following: Let S_1 and S_2 be slowly decreasing; when can we solve $S_1 * T = W_1$, $S_2 * T = W_2$, where $W_1 W_2 \in \mathfrak{D}'$ are given? Clearly a necessary condition is $S_1 * W_2 = S_2 * W_1$. But, even for differential operators this condition cannot be sufficient, for J_1 and J_2 may not be relatively prime. If we assume J_1, J_2 relatively prime, then is $S_1 * W_2 = S_2 * W_1$ sufficient? Since S_1 is invertible, we can reduce the problem to the case $W_1 = 0$. The problem becomes: If $S_1 * W_2 = 0$, can we find a T such that $S_2 * T = W_2$ and $S_1 * T = 0$? This suggests that we try to apply the methods of this paper to the subspace of $\mathfrak{D}'$ which is the kernel of S_1. For this, we need a "good" description of the Fourier transform of the dual of this kernel which is $\mathfrak{D}/S_1 * \mathfrak{D}$. This is just Problem 2.

In this connection we should add that the questions of simultaneous C^∞ (entire) ellipticity in all variables for a system of differential operators can be reduced to the case of a single differential operator as has been shown by the work of L. Hörmander ("Differentiability properties of solutions of system of differential equations," *Arkiv för Matematik*, vol. 3 (1958), pp. 527-535, and C. Lech, "A metric property of the zeros of a complex polynomial ideau," *Arkiv för Matematik*, vol. 3 (1958), pp. 543-554.)

However, the method of Hörmander and Lech fails in the case of systems of convolution equations. For, we can easily construct $(n=1)$ $J_1, J_2 \in \boldsymbol{E}'$ whose common zeros are infinite in number and lie outside of some angle containing R. Thus, by a result of Schwartz (see [27]) every $T \in \mathfrak{D}'$ which satisfies $S_1 * T = 0$, $S_2 * T = 0$ must be an entire function. But, by Carleman's theorem (see [29]) there is no J in $\boldsymbol{E}'$ which has infinitely many zeros all of which lie outside of some angle containing R. Thus, we cannot reduce the above problem for ideals to the problem for a single $S \in \mathcal{E}'$.

Moreover, I do not know if the methods of Hörmander and Lech can be extended to the problem of ellipticity in x_1.

Appendix, added in proof. We indicate some of the progress made since this paper was written: (The numbers correspond to the numbers used in Section 6.)

1. We can produce an example of a J which is slowly decreasing ($n=1$) but such that the orders of the zeros of J are unbounded.

2. In case J is slowly decreasing we can give ($n=1$) necessary and sufficient conditions in order that E'/JE' should be isomorphic with the space of slowly increasing sequences on the zeros of J. In case J has multiple zeros, a similar result is possible.

4. In case $n>1$ I can produce an example to show that $\mathfrak{D} * \mathfrak{D} \neq \mathfrak{D}$.

7. The study of these lacunary series leads to an extension of the Fabry gap theorem to $n>1$.

9. The results on partial ellipticity can be extended completely to systems of partial differential equations and to a very few other convolution systems (see [15], [16]).

YESHIVA UNIVERSITY.

REFERENCES.

[1] S. Banach, *Théorie des Operations Linéares*, Warsaw, 1932.

[2] A. Beurling, *Lectures on Quasi-Analytic Functions*, unpublished lectures given at the Institute for Advanced Study, 1957.

[3] T. Carleman, *Les Fonctions Quasi-Analytiques*, Paris, 1926.

[4] J. Dieudonné, "Recent developments in the theory of locally convex vector spaces," *Bulletin of the American Mathematical Society*, vol. 59 (1953), pp. 495-512.

[5] L. Ehrenpreis, "Solutions of some problems of division I," *American Journal of Mathematics*, vol. 76 (1954), pp. 883-903.

[6] ———, "Solutions of some problems of division II," *ibid.*, vol. 77 (1955), pp. 286-292.

[7] ———, "Solutions of some problems of division III," *ibid.*, vol. 78 (1956), pp. 685-715.

[8] ———, "Mean periodic functions I," *ibid.*, vol. 77 (1955), pp. 293-328, also appendix in same journal, vol. 77 (1955), pp. 731-733.

[9] ———, "Completely inversible operators," *Proceedings of the National Academy of Sciences*, vol. 41 (1955), pp. 945-946.

[10] ———, "General theory of elliptic equations," *ibid.*, vol. 42 (1956), pp. 39-41.

[11] ———, "Analytic functions and the Fourier transform of distributions I," *Annals of Mathematics*, vol. 63 (1956), pp. 129-159.

[12] ———, "On the theory of kernels of Schwartz," *Proceedings of the American Mathematical Society*, vol. 7 (1956), pp. 713-718.

[13] ———, "Sheaves and differential equations," *ibid.*, vol. 7 (1956), pp. 1131-1138.

[14] ———, "Theory of infinite derivatives," *American Journal of Mathematics*, vol. 81 (1959), pp. 799-845.

[15] ———, "The fundamental principle for linear constant coefficient partial differential equations and some of its applications," to appear in the *Proceedings of the International Symposium on Linear Spaces*, Jerusalem, 1960.

[16] ———, *Treatise on linear constant coefficient partial differential equations*, to appear.

[17] L. Garding, "Dirichlet's problem for linear elliptic partial differential equations," *Mathematica Scandinavica*, vol. 1 (1953), pp. 237-255.

[18] A. Grothendieck, "Sur les espaces ($\mathcal{F}$) et ($\mathcal{DF}$)," *Summa Basiliensis Mathematicae*, vol. 3 (1954), pp. 57-121.

[19] ———, *Produits Tensoriels Topologiques et Espaces Nucléaires*, Memoirs of the American Mathematical Society, no. 16, 1955.

[20] L. Hörmander, "On the theory of general partial differential operators," *Acta Mathematica*, vol. 94 (1955), pp. 161-248.

[21] B. Malgrange, "Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution," thesis, Paris, 1955, *Annals de l'Institut Fourier* (Grenoble), vol. 6 (1956), pp. 271-255.

[22] R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*, New York, 1934.

[23] I. G. Petrowsky, "Sur l'analyticité des solutions des systèmes d'équations differentielles," *Matematiceskii Sbornik*, vol. 47 (1939), pp. 3-70.

[24] L. Schwartz, *Théorie des Distributions*, vols. I, II, Paris, 1950-51.

[25] ———, *Etude des sommes d'exponentielles réelles*, Paris, 1943.

[26] ———, "Approximation d'une fonction quelconque par des sommes d'exponentielles imaginaires," *Toulouse Université, Annales de la Faculté des Sciences*, vol. 6 (1942), pp. 111-174.

[27] ———, "Théorie général des fonctions moyènne-periodiques," *Annals of Mathematics*, vol. 48 (1947), pp. 857-929.

[28] T. Shirota, "On solutions of partial differential equations with a parameter," *Procedings o fthe Japan Academy*, vol. 32 (1956), pp. 401-405.

[29] E. C. Titchmarsh, *The Theory of Functions*, Oxford, 1932.

[30] L. Wallbroeck, "Le calcul symbolique dans les algèbres commutative," *Journal de Mathematiques Pures et Appliquées*, vol. 33 (1954), pp. 147-186.

HOMOMORPHISMS OF COMMUTATIVE BANACH ALGEBRAS.*[1]

By W. G. Bade and P. C. Curtis, Jr.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be commutative Banach algebras, and ν be an arbitrary (not necessarily continuous) homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$. This paper is a study of continuity properties of ν which arise from the algebraic structure of $\mathfrak{A}$. The main results are grouped around the following topics:

(a) The degree of discontinuity which ν may have on the set of idempotents in $\mathfrak{A}$.

(b) The localization of the discontinuity of ν to a finite set of points of the structure space $\Phi_{\mathfrak{A}}$ of $\mathfrak{A}$ when $\mathfrak{A}$ is a regular algebra in the sense of Silov.

(c) The question of the existence of discontinuous homomorphisms (or equivalently, the existence of incomplete multiplicative norms) on the algebra $C(\Omega)$ of all continuous functions on a compact Hausdorff space.

(d) The construction of algebras which are not normed algebras under any norm.

If $\nu: \mathfrak{A} \rightarrow \mathfrak{B}$, then the function $|x| = \|\nu(x)\|$, $x \in \mathfrak{A}$, is a multiplicative semi-norm on $\mathfrak{A}$. Conversely (cf. Section 1), every multiplicative semi-norm is the norm of a homomorphism. Thus all our results on continuity of homomorphisms could be stated equivalently in terms of continuity properties of multiplicative semi-norms on $\mathfrak{A}$. We have chosen the homomorphism approach as it reveals the methods more clearly.

Section 1 contains preliminary material concerning adjunction of units and the relation of multiplicative semi-norms to homomorphisms. The first theorem of Section 2 is the key result of the paper: If a bounded sequence $\{g_n\}$ of elements of $\mathfrak{A}$ is separated by orthogonal relative units (elements h_n of $\mathfrak{A}$ satisfying $g_n h_n = g_n$, $h_n h_m = 0$, $m \neq n$), then under any homomorphism

* Received November 10, 1959.

[1] This work was supported by the Office of Naval Research under contract Nonr233-(59), the Air Force Office of Scientific Research under contract SAR/AF-49(636)-153 and by the National Science Foundation.

$\nu: \mathfrak{A} \to \mathfrak{B}$, the norms of the elements $\nu(g_n)$ in $\mathfrak{B}$ cannot grow faster than the norms of the relative units h_n in $\mathfrak{A}$. This result is best possible. From this theorem we obtain an important property of ν on the set $\mathfrak{P}$ of idempotents of $\mathfrak{A}$: there exists a constant M such that

$$\| \nu(p) \| \leqq M \| p \|^2, \; p \in \mathfrak{P}.$$

Again the result is sharp. Thus if $\mathfrak{P}$ is a bounded set in $\mathfrak{A}$, it remains bounded under any homomorphism.

In Section 3 the algebra $\mathfrak{A}$ is supposed to be semi-simple, with unit, and to be regular in the sense of Silov. Regarding $\mathfrak{A}$ as an algebra of continuous functions on its structure space $\Phi_{\mathfrak{A}}$, it is shown that if $\nu: \mathfrak{A} \to \mathfrak{B}$, then there is a finite set F of points of $\Phi_{\mathfrak{A}}$ such that ν is continuous on the ideal of functions in $\mathfrak{A}$ which vanish identically in any given neighborhood of F. Examples are given where the set F is not empty.

The results of Section 3 are considerably strengthened in Section 4 for the case of the special Banach algebra $C(\Omega)$. Any homomorphism $\nu: C(\Omega) \to \mathfrak{B}$ has a decomposition $\nu = \mu + \lambda$, where μ is a continuous homomorphism of $C(\Omega)$, coinciding with ν on a dense subalgebra, and λ maps into the radical of $\mathfrak{B}$. Moreover,

$$\nu\overline{(C(\Omega))} = \mu(C(\Omega)) \oplus \lambda\overline{(C(\Omega))},$$

the direct sum being topological. If the radical of $\mathfrak{B}$ is nil, then $\lambda \equiv 0$ and ν is therefore continuous. It is shown that the existence of an incomplete multiplicative norm on $C(\Omega)$, or of a discontinuous homomorphism, is equivalent to the existence of a non-trivial homomorphism of some maximal ideal of $C(\Omega)$ into a radical Banach algebra. We know of no example of such a homomorphism.

Section 5 deals with the non normability of certain quotient algebras. Explicitly, if $\mathfrak{A}$ is a semi-simple regular algebra with unit, $\varphi_0 \in \Phi_{\mathfrak{A}}$, and $\mathfrak{J}(\varphi_0)$ is the ideal of all functions in $\mathfrak{A}$ which vanish in a neighborhood of φ_0, then the algebra $\mathfrak{A}/\mathfrak{J}(\varphi_0)$ is not normable whenever φ_0 is the limit of a sequence in $\Phi_{\mathfrak{A}}$. Section 6 contains a discussion of a Banach algebra due to C. Feldman [2] which shows that the theorems of Section 1 cannot be improved. It also provides an example of an algebra with one-dimensional radical which admits two inequivalent complete multiplicative norms. This shows that the theorem of Gelfand [3] to the effect that semi-simple commutative Banach algebras have unique Banach algebra topologies cannot be generalized even to algebras with finite dimensional radical.

Our work is related in spirit to the important results of Gelfand [3],

Rickart [8], Silov [9, Theorem 8], Yood [11], and others on continuity of homomorphisms and uniqueness of norms for Banach algebras. However, the methods and results are very different. Typical results of their work are the theorem of Rickart that any homomorphism of a Banach algebra into a commutative semi-simple Banach algebra is continuous. In these results essential use is made of completeness and semi-simplicitly of the range of the homomorphism. In our work the special assumptions are placed on the domain algebra, the range being arbitrary. The conjecture [2] that on $C(\Omega)$ every multiplicative norm $\|\cdot\|$ is complete and equivalent to the supremum norm arises naturally from a theorem of Kaplansky [5], that one necessarily has $\| x \| \geqq \sup_{\Omega} | x(\omega) |$, $x \in C(\Omega)$.

1. Preliminaries. Let $\mathfrak{A}$ be a commutative Banach algebra and ν be a homomorphism of $\mathfrak{A}$ into a Banach algebra $\mathfrak{B}$. Before beginning the study of general properties of ν, it is convenient for technical reasons to make some simplifications. First, by confining attention to $\overline{\nu(\mathfrak{A})}$ we may always assume that ν maps $\mathfrak{A}$ into a commutative Banach algebra $\mathfrak{B}$. Similarly, if $\mathfrak{A}$ has a unit e, we may assume that $\mathfrak{B}$ has a unit e' and that $\nu(e) = e'$. If $\mathfrak{A}$ does not have a unit, then we may form $\mathfrak{A}' = \mathfrak{A} \oplus \{\lambda e\}$ in the usual way. If $\mathfrak{B}$ has a unit, let $\mathfrak{B}' = \mathfrak{B}$; otherwise let $\mathfrak{B}' = \mathfrak{B} \oplus \{\lambda e'\}$. For $x' \in \mathfrak{A}'$, $\| x' \| = \| x \| + | \lambda |$, where $x' = x + \lambda e$, $x \in \mathfrak{A}$. If e' is adjoined to $\mathfrak{B}$ to form $\mathfrak{B}'$, then $\mathfrak{B}'$ is normed in the same way. In either case it is clear that any homomorphism ν of $\mathfrak{A}$ into $\mathfrak{B}$ may be extended to a homomorphism ν' of $\mathfrak{A}'$ into $\mathfrak{B}'$ and that continuity properties of ν are uneffected by this extension. In light of these remarks we shall always assume that $\mathfrak{A}$ and $\mathfrak{B}$ are commutative Banach algebras with units e, e' respectively and that $\nu(e) = e'$.

The study of homomorphisms ν of a Banach algebra $\mathfrak{A}$, commutative or not, is equivalent to the study of multiplicative semi-norms as the following makes clear.

1.1. *Definition.* Let $\mathfrak{A}$ be a Banach algebra with unit. A *multiplicative semi-norm* on $\mathfrak{A}$ is a function $|\cdot|$ on $\mathfrak{A}$ to $[0,\infty)$ satisfying

(i) $|x+y| \leqq |x| + |y|$, $x, y \in \mathfrak{A}$,

(ii) $|xy| \leqq |x|\,|y|$, $x, y \in \mathfrak{A}$,

(iii) $|\alpha x| = |\alpha|\,|x|$, $x \in \mathfrak{A}$, α scalar,

(iv) $|e| = 1$.

[2] An early paper of Gelfand and Naimark contains a statement [4, Lemma 2] which, with Kaplansky's theorem, would imply this conjecture. However, the proof contains a serious gap.

If in addition $|x| = 0$ implies $x = 0$, then $|\cdot|$ will be called a *multiplicative norm.*

Clearly, for a multiplicative semi-norm,

$$||x| - |y|| \leqq |x - y|, \text{ hence } |x| = |y| \text{ if } |x - y| = 0.$$

1.2. THEOREM. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be Branch algebras with units e, e' respectively. If ν is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ with $\nu(e) = e'$, then the function $|x| = \|\nu(x)\|$, $x \in \mathfrak{A}$, is a multiplicative semi-norm on $\mathfrak{A}$. Conversely, if $|\cdot|$ is a multiplicative semi-norm on $\mathfrak{A}$, then there exists ν of $\mathfrak{A}$ into a Banach algebra $\mathfrak{B}$ such that $|x| = \|\nu(x)\|$, $x \in \mathfrak{A}$.*

Proof. The first assertion is clear. To prove the second we note that the set $\mathfrak{R} = \{x \mid |x| = 0\}$ is an ideal and $|\cdot|$ is constant on the cosets $[x + \mathfrak{R}]$ of $\mathfrak{A}/\mathfrak{R}$. Thus $\mathfrak{A}/\mathfrak{R}$ is a normed algebra under the norm $|x + \mathfrak{R}| = |x|$. Let ν be the natural homomorphism of $\mathfrak{A}$ into the completion of $\mathfrak{A}/\mathfrak{R}$ in this norm. It has the required properties.

2. The main boundedness theorem.

The first theorem of this section is the main technical device of the paper. We are indebted to Y. Katznelson for suggestions which greatly shortened the original proof.

2.1. THEOREM. *Let $\mathfrak{A}$ be a commutative Banach algebra and ν be a homomorphism of $\mathfrak{A}$ into a Banach algebra $\mathfrak{B}$. If $\{g_n\}$ and $\{h_n\}$ are sequences from $\mathfrak{A}$ satisfying*

(i) $g_n h_n = g_n$, $n = 1, 2, \cdots$,

and

(ii) $h_m h_n = 0$, $m \neq n$,

then

$$\sup \|\nu(g_n)\| / \|g_n\| \, \|h_n\| < \infty.$$

Proof. Suppose on the contrary that

$$\limsup \|\nu(g_n)\| / \|g_n\| \|h_n\| = +\infty.$$

We may suppose $\|g_n\| = 1$, $n = 1, 2, \cdots$. Clearly $\|h_n\| \geqq 1$ by (i). It will be shown that a suitable linear combination of the elements h_n must map into an element of infinite norm. Select distinct elements q_{ij}, $i, j = 1, 2, \cdots$, from the sequence $\{g_n\}$ such that

$$(*) \qquad \|\nu(q_{ij})\| \geqq 4^{i+j} \|p_{ij}\|, \qquad i, j = 1, 2, \cdots,$$

where p_{ij} is the relative unit h_m corresponding to $g_m = q_{ij}$. Define

$$f_i = \sum_{j=1}^{\infty} q_{ij}/2^j, \qquad i = 1, 2, \cdots.$$

The equation $p_{ij}f_i = 2^{-j}q_{ij}$ and (*) show $\nu(f_i) \neq 0$. For each integer i select an integer j_i so large that $2^{j_i} > \| \nu(f_i) \|$ and define

$$y = \sum_{i=1}^{\infty} p_{ij_i}/2^i \| p_{ij_i} \|.$$

It follows from (i) that

$$f_i y = q_{ij_i}/2^{(i+j_i)} \| p_{ij_i} \|, \qquad i = 1, 2, \cdots.$$

Thus, using (*),

$$\| \nu(y) \| \, \| \nu(f_i) \| \geqq \| \nu(f_i y) \| \geqq 2^{(i+j_i)} > 2^i \| \nu(f_i) \|.$$

Thus $\| \nu(y) \| > 2^i$ for every integer i.

2.2. COROLLARY. *Let ν be an arbitrary homomorphism of the commutative Banach algebra $\mathfrak{A}$ into a Banach algebra $\mathfrak{B}$. If $\{p_n\}$ is a sequence of orthogonal idempotents in $\mathfrak{A}$, i.e. $p_m p_n = 0$ for $m \neq n$, then there exists a constant M such that*

$$\| \nu(p_n) \| \leqq M \| p_n \|^2, \qquad n = 1, 2, \cdots.$$

Proof. The result follows by taking $g_n = h_n = p_n$ in Theorem 2.1.

The next theorem shows that the constant M may be chosen independent of the sequence $\{p_n\}$.

2.3. THEOREM. *Let ν be a homomorphism of the commutative Banach algebra $\mathfrak{A}$ into a Banach algebra $\mathfrak{B}$ and let $\mathfrak{P}$ denote the set of idempotents in $\mathfrak{A}$. There exists a constant M such that*

$$\| \nu(p) \| \leqq M \| p \|^2, \qquad p \in \mathfrak{P}.$$

If $\mathfrak{P}$ is a bounded set in $\mathfrak{A}$, its image under any homomorphism is bounded.

Proof. By the remarks of Section 1 we may suppose $\mathfrak{A}$ has a unit e and that $\nu(e)$ is the unit of $\mathfrak{B}$. Supposing the theorem false, we shall construct an orthogonal sequence which contradicts Corollary 2.2.

Let $\mathfrak{P}_1$ denote the set of $p \in \mathfrak{P}$ such that

$$\sup_{q \leqq p} \| \nu(q) \| / \| q \|^2 = +\infty.$$

By assumption $e \in \mathfrak{P}_1$. Note that if $p \in \mathfrak{P}_1$ and $q \leqq p$, then either q or $p - q$ is in $\mathfrak{P}_1$. For otherwise, there is a constant K such that $\| \nu(r) \| \leqq K \| r \|^2$

for all $r \leqq q$ and $r \leqq p-q$. If $s \leqq p$, we may write $s = sq + s(p-q)$, so

$$\| \nu(s) \| \leqq K[\| sq \|^2 + \| s(p-q) \|^2] \leqq K \| s \|^2 [\| q \|^2 + \| p-q \|^2],$$

contradicting the assumption that p belongs to $\mathfrak{P}_1$.

For purposes of an induction let r_1 belong to $\mathfrak{P}_1$ and choose $q_1 \leqq r_1$ such that

$$\| \nu(q_1) \| / \| q_1 \|^2 > 16 \| r_1 \|^4 [2 + 2 \| \nu(r_1) \| / \| r_1 \|^2].$$

Then

$$\begin{aligned} \| \nu(r_1 - q_1) \| / \| r_1 - q_1 \|^2 &> [\| \nu(q_1) \| - \| \nu(r_1) \|] / [\| r_1 \| \| q_1 \| + \| q_1 \|]^2 \\ &\geqq [\| \nu(q_1) \| / 4 \| r_1 \|^2 \| q_1 \|^2] - [\| \nu(r_1) \| / \| r_1 \|^2] \\ &> 4 \| r_1 \|^2 [2 + \| \nu(r_1) \| / \| r_1 \|^2]. \end{aligned}$$

Let r_2 be the member of the pair $q_1, r_1 - q_1$, which is in $\mathfrak{P}_1$. Then clearly,

$$\| \nu(r_2) \| / \| r_2 \|^2 > 4 \| r_1 \|^2 [2 + \| \nu(r_1) \| / \| r_1 \|^2].$$

By an exactly similar arguments we obtain inductively a sequence $\{r_k\}$ of idempotents in $\mathfrak{P}_1$ such that $r_{k+1} \leqq r_k$ and

$$\| \nu(r_k) \| / \| r_k \|^2 > 4 \| r_{k-1} \|^2 [k + \| \nu(r_{k-1}) \| / \| r_{k-1} \|^2], \qquad k = 2, 3, \cdots.$$

Define $p_k = r_k - r_{k+1}$. Then $p_k p_l = 0$ if $k \neq l$ and

$$\begin{aligned} \| \nu(p_k) \| / \| p_k \|^2 &\geqq [\| \nu(r_{k+1}) \| / 4 \| r_k \|^2 \| r_{k+1} \|^2] - [\| \nu(r_k) \| / \| r_k \|^2] \\ &> k + 1, \qquad\qquad k = 2, 3, \cdots, \end{aligned}$$

contradicting Corollary 2.2.

3. Homomorphisms of regular algebras. In this section $\mathfrak{A}$ will be a commutative semi-simple Banach algebra with unit which is regular in the sense of Silov [9]. We shall regard $\mathfrak{A}$ via the Gelfand isomorphism as an algebra of continuous functions on its structure sapce $\Phi_{\mathfrak{A}}$.[3] Recall that the property of being regular is equivalent to the condition that given any two disjoint closed sets F_1 and F_2 in $\Phi_{\mathfrak{A}}$, there exists a function in $\mathfrak{A}$ which is zero on F_1 and one on F_2 (cf. [9] or Loomis [7, p. 84]). We shall show that if ν is any homomorphism of $\mathfrak{A}$ into a Banach algebra, there exists a finite set F of points of $\Phi_{\mathfrak{A}}$ such that for any neighborhood V of F the restriction of ν to the ideal $\mathfrak{J}(V) = \{f \in \mathfrak{A} \mid f(V) = 0\}$ is continuous. We shall also obtain information as to how the norm of ν on $\mathfrak{J}(V)$ depends on the neighborhood V.

[3] The topology of $\mathfrak{A}$ is always the Banach algebra topology of $\mathfrak{A}$, rather than the relative sup norm topology, except, of course, when they coincide.

3.1. *Definition.* We denote by $\mathfrak{G}$ the family of all open sets $E \subseteq \Phi_{\mathfrak{A}}$ with the property that

$$\sup \| \nu(g) \| / \| g \| \, \| h \| = M_E < \infty$$

for all functions g and h having carriers in E and such that $gh = g$.

We shall show that $\mathfrak{G}$ contains a maximal open set whose complement is finite. This will be accomplished through a sequence of lemmas. The first of these, a direct consequence of the main boundedness theorem of Section 2, shows the existence of many open sets in $\mathfrak{G}$.

3.2. Lemma. *If $\{E_n\}$ is any sequence of disjoint open sets in $\Phi_{\mathfrak{A}}$, then $E_n \in \mathfrak{G}$ for all sufficiently large n.*

Proof. If the lemma is false there exists an infinite sequence $\{E_m\}$ of disjoint open sets and functions g_m, h_m in $\mathfrak{A}$, whose carriers lie in E_m such that

(i) $\| g_m \| = 1$,

(ii) $g_m h_m = g_m$,

and

(iii) $\| \nu(g_m) \| > m \| h_m \|$,

which contradicts Theorem 2.1.

The next task is to prove that $\mathfrak{G}$ is closed under arbitrary unions. Several lemmas will be required.

3.3. Lemma. *Let E_1 and E_2 belong to $\mathfrak{G}$. If G is an open set such that $\bar{G} \subseteq E_2$, then $E_1 \cup G$ is in $\mathfrak{G}$.*

Proof. By regularity we can choose a function $u_1 \in \mathfrak{A}$ which is one on a neighborhood of E_2' and zero on a neighborhood of $\bar{G}$. Let $u_2 = 1 - u_1$. Since $\operatorname{car}(u_1) \cap \bar{G} = \phi$ and $\operatorname{car}(u_2) \cap E_2' = \phi$ we can find functions v_1 and v_2 in $\mathfrak{A}$ such that

$$u_1 v_1 = u_1, \qquad \operatorname{car}(v_1) \cap \bar{G} = \phi,$$

$$u_2 v_2 = u_2, \qquad \operatorname{car}(v_2) \cap E_2' = \phi.$$

Let $H = E_1 \cup G$ and suppose $\operatorname{car}(g) \subseteq H$, $\operatorname{car}(h) \subseteq H$ and $gh = g$. Then

$$\operatorname{car}(gu_i) \subseteq E_i, \qquad \operatorname{car}(hv_i) \subseteq E_i, \qquad i = 1, 2,$$

and

$$gu_i = ghu_i v_i = (gu_i)(hv_i), \qquad i = 1, 2.$$

Since E_1 and E_2 belong to $\mathfrak{G}$,

$$\begin{aligned}\| \nu(g) \| &\leqq \| \nu(gu_1) \| + \| \nu(gu_2) \| \\ &\leqq M_{B_1} \| gu_1 \| \| hv_1 \| + M_{B_2} \| gu_2 \| \| hv_2 \| \\ &\leqq \{M_{B_1} \| u_1 \| \| v_1 \| + M_{B_2} \| u_2 \| \| v_2 \|\} \| g \| \| h \|,\end{aligned}$$

showing H belongs to $\mathfrak{G}$.

3.4. COROLLARY. *If $E_1, E_2 \in \mathfrak{G}$ and G is open with $\bar{G} \subseteq E_1 \cup E_2$, then $G \in \mathfrak{G}$.*

Proof. The closed set $F = E_1' \cap \bar{G}$ is contained in E_2. Let U be open with $F \subseteq U \subseteq \bar{U} \subseteq E_2$. Then $G \subseteq E_1 \cup U$ which belongs to $\mathfrak{G}$ by Lemma 3.3. Thus $G \in \mathfrak{G}$.

3.5. LEMMA. *If $E_1, E_2 \in \mathfrak{G}$, then $E_1 \cup E_2 \in \mathfrak{G}$.*

Proof. Suppose $E_1 \cup E_2 \notin \mathfrak{G}$. We note that if F is closed and $F \subseteq E_1 \cup E_2$, then $G = (E_1 \cup E_2) \sim F$ is also not in $\mathfrak{G}$. For choose open sets U and V such that

$$F \subseteq V \subseteq \bar{V} \subseteq U \subseteq \bar{U} \subseteq E_1 \cup E_2.$$

Then $U \in \mathfrak{G}$ by Corollary 3.4. If $G \in \mathfrak{G}$, then by Lemma 3.3, $E_1 \cup E_2 = G \cup V \in \mathfrak{G}$, contrary to assumption.

Now if $E_1 \cup E_2 \notin \mathfrak{G}$, we can find g_1, h_1, such that $g_1 = g_1 h_1$, $\operatorname{car}(h_1) \subseteq E_1 \cup E_2$ and

$$\| \nu(g_1) \| > \| g_1 \| \| h_1 \|.$$

Pick U_1 open such that $\operatorname{car}(h_1) \subseteq U_1 \subseteq \bar{U}_1 \subseteq E_1 \cup E_2$. Then by the remark above $G_2 = (E_1 \cup E_2) \sim \bar{U}_1 \notin \mathfrak{G}$, and hence there exist $g_2, h_2, g_2 h_2 = g_2$, and $\operatorname{car}(h_2) \subseteq G_2$, satisfying

$$\| \nu(g_2) \| \geqq 2 \| g_2 \| \| h_2 \|.$$

Continuing inductively we obtain sequences $\{g_n\}$, $\{h_n\}$ such that

(1) $g_n h_n = g_n$,

(2) $\| \nu(g_n) \| \geqq n \| g_n \| \| h_n \|$,

(3) the carriers of the h_n lie in disjoint open sets. This contradiction of Theorem 2.1 completes the proof.

3.6. COROLLARY. *$\mathfrak{G}$ is closed under arbitrary unions.*

Proof. Let $E_0 = \cup E_\alpha$, where $E_\alpha \in \mathfrak{G}$. Suppose $E_0 \notin \mathfrak{G}$. Then if F is closed and $F \subseteq E_0$, we note $E_0 \sim F \notin \mathfrak{G}$. For by compactness F is covered by a finite union E_1 of sets in $\mathfrak{G}$, i.e. a set in $\mathfrak{G}$. Thence $E_0 = (E_0 \sim F)$

$\cup E_1 \in \mathfrak{G}$. Now a repetition of the construction of the last proof yields a contradiction.

3.7. THEOREM. *Let $\mathfrak{A}$ be a commutative semi-simple regular Banach algebra with unit and let ν be an arbitrary homomorphism of $\mathfrak{A}$ into a Banach algebra. There exists a finite set F of points of $\Phi_{\mathfrak{A}}$ and a constant M such that*

$$\| \nu(g) \| \leqq M \| g \| \| h \|$$

for all functions g and h in $\mathfrak{A}$ having carriers in $\Phi_{\mathfrak{A}} \sim F$ and such that $gh = g$.

Proof. By Corollary 3.6 the class $\mathfrak{G}$ contains a maximal open set G_0. Let F be its complement. If F is infinite, we may separate a sequence of its elements by disjoint open sets E_n. By Lemma 3.2 $E_n \in \mathfrak{G}$ for large n. Thus G_0 must contain points of its complement. This contradiction shows F is finite.

3.8. *Definition.* The finite set F of Theorem 3.7 will be called the *singularity set* of ν.

If V is an open set in $\Phi_{\mathfrak{A}}$ we write $\mathfrak{J}(V) = \{f \in \mathfrak{A} \mid f(V) = 0\}$.

3.9. COROLLARY. *If V is any neighborhood of the singularity set F of ν, then the restriction of ν to $\mathfrak{J}(V)$ is continuous, and*

$$\| \nu(f) \| \leqq M \| f \| \| h \|, \qquad f \in \mathfrak{J}(V),$$

where h is any function in $\mathfrak{A}$ which is one on V' and which vanishes in a neighborhood of F.

We conclude this section with some examples of discontinuous isomorphisms of regular algebras, to show that the singuarity set F is not always empty. Another example will be given in Section 6. Discontinuous isomorphisms can always be constructed when there is a maximal ideal $\mathfrak{M}$ in $\mathfrak{A}$ such that $\mathfrak{M}^2$ is not closed in $\mathfrak{A}$. It is an open question whether such isomorphisms can be constructed in algebras such as $C(\Omega)$ or the group algebra of a locally compact abelian group (cf. [1]) where $\mathfrak{M}^2 = \mathfrak{M}$ for every regular maximal ideal.

Example 1. Let $\mathfrak{A}$ be the algebra l_p, $1 \leqq p < \infty$, under pointwise multiplication. (There is no unit, but we can adjoin one and consider l_p as a maximal ideal in $\mathfrak{A} \oplus \{\lambda e\}$.) It is easy to see that $(l_p)^2 = l_{p/2}$, which is a proper dense subset of l_p and thus cannot be closed. Let θ be a discontinuous linear functional in l_p which vanishes on $(l_p)^2$. Define $\mathfrak{B} = l_p \oplus \{\lambda r\}$, where $r(l_p) = 0 = r^2$, and define

$$\nu(x) = x + \theta(x)r, \qquad x \in l_p.$$

Then ν is a discontinuous isomorphism of l_p into $\mathfrak{B}$.

Example 2. Consider the algebra $\mathfrak{D}^1$ of continuously differentiable functions on $[0,1]$ with $\| x \| = \sup | x(t) | + \sup | x'(t) |$. The structure space is $[0,1]$. Define

$$\mathfrak{M} = \{x \mid x(0) = 0\}, \qquad \mathfrak{N} = \{x \mid x(0) = x'(0) = 0\}.$$

If $y \in \mathfrak{N}$, and p_n is any sequence of polynomials in $\mathfrak{M}$ converging uniformly to y', then the polynomials $q_n(t) = \int_0^t p_n(s)\,ds$ lie in $\mathfrak{M}^2$ and converge to y in $\mathfrak{D}^1$. Thus $\mathfrak{M}^2$ is dense in $\mathfrak{N}$. However, $\mathfrak{M}^2$ is not closed since it is easily seen that $y''(0)$ exists for every y in $\mathfrak{M}^2$. We may now use the argument of the last example to construct a discontinuous isomorphism of $\mathfrak{M}$ (and hence of $\mathfrak{D}^1$) into a Banach algebra.

4. Homomorphisms of $C(\Omega)$. In this section we consider the case that $\mathfrak{A}$ is the algebra $C(\Omega)$ of all continuous real or complex functions on a compact Hausdorff space Ω. Since $\mathfrak{M}^2 = \mathfrak{M}$ for every maximal ideal, the techniques for constructing discontinuous homomorphisms of the last section fail. In fact there are no known examples of discontinuous homomorphisms of $C(\Omega)$ for any Ω. In this section we shall strengthen the main result (Theorem 3.7) of the last section and exhibit the precise role the radical of the image must play if the homomorphism is to be discontinuous. We shall see that in the case of $C(\Omega)$ there is a decomposition of the homomorphism into the sum of two mappings, one continuous, and the other mapping into the radical of $\mathfrak{B}$. If the radical of the image is nil, the "radical" part of the homomorphism ν must be trivial, forcing ν to be continuous.

As before we denote by $F = \{\omega_1, \cdots, \omega_n\}$ the finite singularity set for $\nu\colon C(\Omega) \to \mathfrak{B}$. It is convenient to introduce the following classes of functions:

1. $\mathfrak{M}(F)$ is the intersection of the n maximal ideals $\mathfrak{M}(\omega_i)$, $\omega_i \in F$.

2. $\mathfrak{J}(F)$ is the ideal of functions each of which vanishes in a neighborhood of F, the neighborhood depending on the function.

3. $\mathfrak{K}(F)$ is the dense subalgebra of $C(\Omega)$ consisting of those functions f such that $f(\omega) \equiv f(\omega_i)$ in a neighborhood of each point $\omega_i \in F$, the neighborhoods varying with f.

In the algebra $C(\Omega)$ relative units may always be chosen to have norm one. This fact allows a significant strengthening of Theorem 3.7.

4.1. THEOREM. *Let ν be a homomorphism of $C(\Omega)$ into a Banach algebra. If F denotes the singularity set of ν, then ν is continuous on the dense subalgebra $\mathfrak{K}(F)$ of $C(\Omega)$.*

Proof. It follows from Theorem 3.7 and the remark above that there is a constant M such that

$$\| \nu(g) \| \leqq M \| g \|, \qquad g \in \mathfrak{J}(F).$$

We now select functions e_i, $0 \leqq e_i \leqq 1$, such that $e_i e_j = 0$, $i \neq j$, and $e_i(\omega) \equiv 1$ in a neighborhood of $\omega_i \in F$. Then for any $f \in \mathfrak{K}(F)$, $f - \sum_{i=1}^{n} f(\omega_i) e_i \in \mathfrak{J}(F)$, so

$$\begin{aligned} \| \nu(f) \| &\leqq \| \nu(f - \sum_{i=1}^{n} f(\omega_i) e_i) \| + \| \sum_{i=1}^{n} f(\omega_i) \nu(e_i) \| \\ &\leqq M \| f - \sum_{i=1}^{n} f(\omega_i) e_i \| + \| f \| \sum_{i=1}^{n} \| \nu(e_i) \|. \\ &\leqq [(n+1)M + \sum_{i=1}^{n} \| \nu(e_i) \|] \| f \|, \qquad f \in \mathfrak{K}(F). \end{aligned}$$

Since ν is continuous on $\mathfrak{K}(F)$, it has a unique continuous extension to all of $C(\Omega)$,

4.2. *Definition.* We denote by μ the unique continuous homomorphism of $C(\Omega)$ into $\mathfrak{B}$ which agrees with ν on the dense subalgebra $\mathfrak{K}(F)$. Define

$$\lambda(f) = \nu(f) - \mu(f), \qquad f \in C(\Omega).$$

The mappings μ and λ will be called the *continuous* and *singular* parts of ν. We reserve the letter M now for a constant such that

$$\| \mu(f) \| \leqq M \| f \|, \qquad f \in C(\Omega).$$

The next theorem describes the structure of an arbitrary homomorphism of $C(\Omega)$.

4.3. THEOREM. *Let ν be a homomorphism of $C(\Omega)$ into a commutative Banach algebra $\mathfrak{B}$ and let $\mathfrak{R}$ be the radical of $\overline{\nu(C(\Omega))}$. Let $F = \{\omega_1, \cdots, \omega_n\}$ be the singularity set for ν and μ and λ be the continuous and singular parts of ν. Then:*

(a) *The range of μ is closed in $\mathfrak{B}$ and*

$$\overline{\nu(C(\Omega))} = \mu(C(\Omega)) \oplus \mathfrak{R},$$

the direct sum being topological.

(b) $\mathfrak{R} = \overline{\lambda(C(\Omega))}$.

(c) $\mathfrak{R}\cdot\mu(\mathfrak{M}(F))=0$, *and the restriction of* λ *to* $\mathfrak{M}(F)$ *is a homomorphism.*

(d) *There exist linear transformations* λ_i, $i=1,\cdots,n$, *such that*

(i) $\lambda=\sum_{i=1}^{n}\lambda_i$,

(ii) $\mathfrak{R}=\mathfrak{R}_1\oplus\cdots\oplus\mathfrak{R}_n$,

where $\mathfrak{R}_i=\overline{\lambda_i(C(\Omega))}$, *the direct sum being topological.*

(iii) $\mathfrak{R}_i\cdot\mathfrak{R}_j=0$, $i\neq j$, *and* $\mathfrak{R}_i\cdot\mu(\mathfrak{M}(\omega_i))=0$,

$i=1,\cdots,n$.

(iv) *The restriction of* λ_i *to* $\mathfrak{M}(\omega_i)$ *is a homomorphism.*

Proof. Let $\mathfrak{K}=\{x\mid\mu(x)=0\}$. Since μ is continuous, $\mathfrak{K}$ is a closed ideal in $C(\Omega)$, so there exists a closed set $G\subseteq\Omega$ such that

$$\mathfrak{K}=\{x\mid x(\omega)=0,\omega\in G\}.$$

If we give $C(\Omega)/\mathfrak{K}$ the norm

$$\|x+\mathfrak{K}\|=\inf\{\|y\|\mid y\in[x+\mathfrak{K}]\},$$

then by a theorem of Stone [10, Theorem 85], $C(\Omega)/\mathfrak{K}$ is isometrically isomorphic with $C(G)$, and

$$\|x+\mathfrak{K}\|=\sup_{\omega\in G}|x(\omega)|.$$

On the other hand, the semi-norm $|x|=\|\mu(x)\|$ is constant on the cosets $[x+\mathfrak{K}]$, so we may norm $C(\Omega)/\mathfrak{K}$ by defining $|x+\mathfrak{K}|=|x|$. By a theorem of Kaplansky [5] $|x+\mathfrak{K}|\geqq\|x+\mathfrak{K}\|$, $x\in C(\Omega)$. However, for any $y\in[x+\mathfrak{K}]$, we have $|x|=|y|\leqq M\|y\|$ since μ is continuous. Thus

$$|x+\mathfrak{K}|\leqq M\inf\{\|y\|\mid x-y\in\mathfrak{K}\}=M\|x+\mathfrak{K}\|,$$

showing the two norms are equivalent on $C(\Omega)/\mathfrak{K}$. To show μ has a closed range, suppose $b_0\in\mathfrak{B}$ and $b_0=\lim\mu(x_n)$. Then

$$\|x_m-x_n+\mathfrak{K}\|\leqq|x_m-x_n|=\|\mu(x_m-x_n)\|\to 0.$$

There exists $x_0\in C(\Omega)$ such that $\|x_0-x_n+\mathfrak{K}\|\to 0$. Thus

$$\|\mu(x_0)-\mu(x_n)\|\leqq M\|x_0-x_n+\mathfrak{K}\|\to 0,$$

showing $b_0=\mu(x_0)$. In particular, $\mu(C(\Omega))$ is algebraically and topologically isomorphic with $C(G)$. Thus, necessarily, $\mu(C(\Omega))\cap\mathfrak{R}=(0)$.

We next prove that $\lambda = \nu - \mu$ maps into $\mathfrak{R}$. If $\phi \in \Phi_{\mathfrak{B}_1}$, $\mathfrak{B}_1 = \overline{\nu(C(\Omega))}$, then the functionals ϕ_ν and ϕ_μ on $C(\Omega)$ defined by $\phi_\nu(x) = \phi(\nu(x))$, $\phi_\mu(x) = \phi(\mu(x))$ are multiplicative, and hence continuous. Since they coincide on the dense subalgebra $\mathfrak{K}(F)$, we have $\phi_\nu = \phi_\mu$. Thus $\phi(\lambda(x)) = 0$, $x \in C(\Omega)$, $\phi \in \Phi_{\mathfrak{B}_1}$.

It follows now that $\overline{\nu(C(\Omega))} \subseteq \mu(C(\Omega)) \oplus \mathfrak{R}$. To complete the proof of (a) it suffices to show $\overline{\nu(C(\Omega))} = \mu(C(\Omega)) \oplus \mathfrak{R}$ since the algebraic direct sum must be topological as both of the factors are closed [6]. If $b = \lim \nu(x_n)$, then since $\mu(C(\Omega))$ is closed in $\mathfrak{B}$,[4]

$$\begin{aligned}\| \nu(x_m - x_n) \| &\geqq \rho_{\mathfrak{B}}(\nu(x_m - x_n)) \\ &= \rho_{\mathfrak{B}}(\mu(x_m - x_n)) \\ &= \rho_{\mu(C(\Omega))}(\mu(x_m - x_n)) \\ &\geqq M^{-1} \| \mu(x_m - x_n) \|.\end{aligned}$$

Thus there exists $x_0 \in C(\Omega)$ such that $\mu(x_0) = \lim \mu(x_n)$. If $r = b - \mu(x_0)$, then $r = \lim \lambda(x_n)$, so $r \in \mathfrak{R}$. This completes the proof of (a). Statement (b) follows from (a) and the last argument.

We next prove $\mathfrak{R} \cdot \mu(\mathfrak{M}(F)) = 0$. Since $\mathfrak{J}(F)$ is dense in $\mathfrak{M}(F)$, it is enough to prove that $\mu(z)\lambda(x) = 0$, $z \in \mathfrak{J}(F)$, $x \in C(\Omega)$. Now $xz \in \mathfrak{J}(F)$ and ν and μ agree on $\mathfrak{J}(F)$. Thus

$$\begin{aligned}\mu(z)\lambda(x) &= \mu(z)[\nu(x) - \mu(x)] \\ &= \nu(z)\nu(x) - \mu(z)\mu(x) \\ &= \nu(zx) - \mu(zx) = 0.\end{aligned}$$

If $x, y \in \mathfrak{M}(F)$, we have

$$\begin{aligned}\lambda(xy) &= \nu(xy) - \mu(xy) \\ &= [\mu(x) + \lambda(x)][\mu(y) + \lambda(y)] - \mu(xy) \\ &= \mu(x)\mu(y) + \lambda(x)\lambda(y) - \mu(xy) \\ &= \lambda(x)\lambda(y)\end{aligned}$$

since the cross product terms vanish. Thus $\lambda \colon \mathfrak{M}(F) \to R$ is a homomorphism and (c) is proved.

For (d) select functions e_i, $i = 1, \cdots, n$, such that e_i is one in a neighborhood of ω_i and $e_i e_j = 0$, $i \neq j$. Define

$$\lambda_i(x) = \lambda(e_i x), \qquad x \in C(\Omega).$$

[4] $\rho_{\mathfrak{A}}(y)$ denotes the spectral radius of y in the algebra $\mathfrak{A}$.

If $x, y \in \mathfrak{M}(\omega_i)$, then $e_i x, e_i y \in \mathfrak{M}(F)$, so

$$\lambda_i(x)\lambda_i(y) - \lambda_i(xy) = \lambda((e_i^2 - e_i)xy) = 0,$$

as $(e_i^2 - e_i)xy \in \mathfrak{J}(F)$. Thus $\lambda_i \colon \mathfrak{M}(\omega_i) \to \mathfrak{R}$ is a homomorphism. That $\mathfrak{R}_i \cdot \mathfrak{R}_j = 0$ is immediate. The relation $\lambda = \sum \lambda_i$ follows from the fact $(1 - \sum e_i)x \in \mathfrak{J}(F)$ for all $x \in C(\Omega)$.

All that remains is to prove (d) (ii) and the fact $\mathfrak{R}_i \cdot \mu(\mathfrak{M}(\omega_i)) = 0$. For these we need the relations

$$(*) \qquad \mu(e_i)\lambda(e_i y) = \lambda(e_i y), \quad y \in C(\Omega),$$

$$(**) \qquad \mu(e_j)\lambda(e_i y) = 0, \quad i \neq j, \quad y \in C(\Omega).$$

For (**) note

$$\begin{aligned} 0 = \nu(e_i)\nu(e_j y) &= \mu(e_i)[\mu(e_j y) + \lambda(e_j y)] \\ &= \mu(e_i)\lambda(e_j). \end{aligned}$$

For (*)

$$\begin{aligned} \lambda(e_i y) &= [\mu(1 - \sum_{j=1}^{n} e_j) + \sum_{j=1}^{n} \mu(e_j)]\lambda(e_i y) \\ &= \mu(e_i)\lambda(e_i y) \end{aligned}$$

by (**) and (c). Now (d) (ii) follows directly. Finally, if $z \in \mathfrak{M}(\omega_i)$,

$$z = \sum e_j z + (1 - \sum e_j)z,$$

the last term being in $\mathfrak{J}(F)$. Thus by (**) and the fact $e_i z \in \mathfrak{M}(F)$, we have

$$\begin{aligned} \lambda(e_i x)\mu(z) &= \lambda(e_i x) \sum_{j=1}^{n} \mu(e_j)\mu(z) \\ &= \lambda(e_i x)\mu(e_i z) = 0. \end{aligned}$$

It is an open problem whether $C(\Omega)$ admits a discontinuous homomorphism. The last theorem allows us to reduce this question to the question of the existence of a homomorphism of a maximal ideal of $C(\Omega)$ into a radical algebra. We summarize the situation in

4.4. THEOREM. *If the algebra $C(\Omega)$ has any one of the following, it has every other.*

(1) *An incomplete multiplicative norm,*

(2) *A discontinuous multiplicative semi-norm,*

(3) *A discontinuous isomorphism into a Banach algebra,*

(4) *A discontinuous homomorphism into a Banach algebra,*

(5) *A homomorphism* λ *into a radical Banach algebra with adjoined unit* $\mathfrak{R} \oplus \{\alpha e\}$, *such that for some maximal ideal* ω_0, $\lambda(\mathfrak{M}(\omega_0)) \subseteq \mathfrak{R}$ *and* $\lambda(\mathfrak{J}(\omega_0)) = 0$.

Proof. We know $(1) \Leftrightarrow (3)$ and $(2) \Leftrightarrow (4)$. Also $(3) \Rightarrow (4) \Rightarrow (5)$, for we may take for λ one of the homomorphisms λ_i of the last theorem. Given (5), the norm $|x| = \| x \| + \| \lambda(x) \|$ defines a multiplicative norm on $\mathfrak{M}(\omega_0)$. This we can raise to $C(\Omega)$, showing $(5) \Rightarrow (1)$.

The next theorem gives a sufficient condition for a homomorphism of $C(\Omega)$ to be continuous.

4.5. THEOREM. *Let* ν *be a homomorphism of* $C(\Omega)$ *into a commutative Banach algebra* $\mathfrak{B}$. *If the radical* $\mathfrak{R}$ *of* $\mathfrak{B}$ *is a nil ideal* (*i.e.* $r^k = 0$ *for some* k *if* $r \in \mathfrak{R}$), *then* ν *is continuous.*

Proof. In view of the splitting $\nu = \mu + \lambda$ it suffices to prove $\lambda \equiv 0$ on $\mathfrak{M}(F)$, since for any $x \in C(\Omega)$,

$$x = [x - \sum_{i=1}^{n} x(\omega_i) e_i] + \sum_{i=1}^{n} x(\omega_i) e_i,$$

where the e_i are orthogonal functions and e_i is identically one on a neighborhood of ω_i. Thus

$$\lambda(x) = \lambda(x - \sum_{i=1}^{n} x(\omega_i) e_i)$$

as the last term belongs to $\mathfrak{K}(F)$. It is enough to show $\lambda(x) = 0$ for $x \geqq 0$, $x \in \mathfrak{M}(F)$. Consider the following functions on the interval $0 < t < 1$:

$$f(t) = e^{-1/t}, \qquad f_n(t) = t^{-n} e^{-1/t}, \qquad g(t) = -[\ln t]^{-1}.$$

They all approach zero as $t \to 0$. The functions $f(x)$, $f_n(x)$, and $g(x)$ are in $\mathfrak{M}(F)$ and $f(x) = x^n f_n(x)$. Since $\lambda(x)^n = 0$ for some n, $\lambda(f(x)) = 0$. However, if $y = g(x)$, then $y \geqq 0$, $y \in \mathfrak{M}(F)$, and hence $\lambda(f(y)) = 0$. But $x = f(y)$. Thus $\lambda(x) = 0$.

5. Non normable algebras. Let $\mathfrak{A}$ be a semi-simple regular algebra with unit, and let $\mathfrak{B}$ be a commutative Banach algebra. For a homomorphism $\nu\colon \mathfrak{A} \to \mathfrak{B}$ with singularity set F we define the ideals $\mathfrak{M}(F)$ and $\mathfrak{J}(F)$ as in Section 4. For any point $\varphi_0 \in \Phi_{\mathfrak{A}}$ the ideal $\mathfrak{J}(\varphi_0)$ is the set of x in $\mathfrak{A}$ each of which vanishes in some neighborhood of φ_0. In this section we investigate the kernel of a *radical homomorphism* of $\mathfrak{A}$, that is, a homomorphism which maps $\mathfrak{M}(F)$ into the radical $\mathfrak{R}$ of $\mathfrak{B}$. It is easy to see that the kernel must contain $\mathfrak{J}(F)$, since if $x \in \mathfrak{J}(F)$, we can find $y \in \mathfrak{J}(F)$ such that $xy = x$. Thus

$$\nu(x) = \nu(x)\nu(y)^n, \qquad n = 1, 2, \cdots,$$

and the right side converges to zero as $\nu(y) \in \mathfrak{R}$. It will be shown that whenever $\overline{\mathfrak{J}(F)} \sim \mathfrak{J}(F)$ contains an orthogonal sequence ($f_m f_n = 0$, $m \neq n$), then the kernel includes elements of $\overline{\mathfrak{J}(F)}$ which are not in $\mathfrak{J}(F)$. This result is used to show that certain quotient algebras of $\mathfrak{A}$ are not normable.

For the first theorem we introduce a condition originally due to Ditkin (cf. [7, p. 86]).

5.1. *Definition.* The algebra $\mathfrak{A}$ satisfies the condition (D) at a point φ_0 in $\Phi_{\mathfrak{A}}$ if for each $x \in \overline{\mathfrak{J}(\varphi_0)}$, there exists a sequence $\{y_n\} \subseteq \mathfrak{J}(\varphi_0)$ such that $\lim_n xy_n = x$.

5.2. LEMMA. *Let ν: $\mathfrak{A} \to \mathfrak{B}$ be a radical homomorphism and $\{f_n\} \subseteq \mathfrak{J}(F)$, $f_m f_n = 0$, $m \neq n$. Then $\nu(f_n)^3 = 0$ for all large n. If $\mathfrak{A}$ satisfies the condition (D) at each point of F, then $\nu(f_n)^2 = 0$ for all large n.*

Proof. Suppose the theorem is false and that $\nu(f_n)^3 \neq 0$ for all n. By normalizing we can assume for convenience that $\| \nu(f_n) \| = \| \nu(f_n)^3 \|$. Choose $z_n \in \mathfrak{J}(F)$ such that

$$\| f_n - z_n \| < 1/n^3 \| f_n \|, \qquad n = 1, 2, \cdots,$$

and define $h_n = nf_n(f_n - z_n)$, $h = \sum_{n=1}^{\infty} h_n$. Then

$$f_n h = f_n h_n = nf_n^3 - nf_n^2 z_n, \qquad n = 1, 2, \cdots,$$

so

$$\nu(f_n)\nu(h) = n\nu(f_n)^3, \qquad n = 1, 2, \cdots,$$

which implies $\| \nu(h) \| \geqq n$, $n = 1, 2, \cdots$. This is the desired contradiction.

If the condition (D) holds at each point of F the argument can be strengthened. For suppose

$$\nu(f_n)^2 \neq 0, \quad \| \nu(f_n) \| = \| \nu(f_n)^2 \|, \qquad n = 1, 2, \cdots.$$

It is easy to see there exist functions $y_n \in \mathfrak{J}(F)$ such that

$$\| f_n - f_n y_n \| < 1/n^3.$$

Let $h_n = n(f_n - f_n y_n)$, $h = \sum h_n$. Then

$$\nu(h)\nu(f_n) = n\nu(f_n^2), \qquad n = 1, 2, \cdots,$$

so $\| \nu(h) \| \geqq n$, $n = 1, 2, \cdots$.

We next examine the case $\mathfrak{A} = C(\Omega)$. If ν is any homomorphism of $C(\Omega)$, then it follows from Theorem 4.1 that ν maps bounded orthogonal

sequences into bounded sequences, whenever the carriers of the members lie in disjoint open sets. We can now remove this last restriction.

5.3. COROLLARY. *Let ν be a homomorphism of $C(\Omega)$ into a Banach algebra and suppose $f_m f_n = 0$, $m \neq n$. Then*

$$\sup \| \nu(f_n) \| / \| f_n \| < \infty.$$

If ν is a radical homomorphism, then $\nu(f_n) = 0$ for all large n.

Proof. It suffices to establish the result for non-negative sequences since the sequences $\{f_n^+\}$ and $\{f_n^-\}$ are orthogonal, where $f_n^+ = f_n \wedge 0, f_n^- = -(f_n \vee 0)$. Thus we may suppose $f_n \geqq 0$. Clearly, all but finitely many elements of the sequence belong to $\mathfrak{M}(F)$, so we may assume $f_n \in \mathfrak{M}(F)$, $n = 1, 2, \cdots$. Let λ be the singular part of ν (cf. Section 4). Applying the second statement of Lemma 5.2 to the sequence $\{f_n^{\frac{1}{3}}\}$, we see $\lambda(f_n) = 0$ for all large n. (The fact that λ is a homomorphism only on $\mathfrak{M}(F)$ does not affect the argument). The result now follows directly.

5.4. THEOREM. *Let $\mathfrak{A}$ be a regular algebra and let $\varphi_0 \in \Phi_{\mathfrak{A}}$. If φ_0 is the limit of a sequence of distinct points $\{\varphi_n\} \subseteq \Phi_{\mathfrak{A}}$, then the algebra $\mathfrak{A}/\mathfrak{J}(\varphi_0)$ is not normable.*

Proof. Suppose $\mathfrak{A}/\mathfrak{J}(\varphi_0)$ is a normed algebra under some norm and let $\nu: \mathfrak{A} \to \mathfrak{A}/\mathfrak{J}(\varphi_0)$ be the natural homomorphism. Since $\mathfrak{J}(\varphi_0)$ is contained in the unique maximal ideal $\mathfrak{M}(\omega_0)$ in $\mathfrak{A}$, $\mathfrak{A}/\mathfrak{J}(\varphi_0)$ has a unique maximal ideal. Thus ν is a radical homomorphism of $\mathfrak{A}$ into the completion of $\mathfrak{A}/\mathfrak{J}(\varphi_0)$, whose kernel is precisely $\mathfrak{J}(\varphi_0)$. The desired contradiction will be obtained if we can show the kernel must be larger. Since $\varphi_0 = \lim \varphi_n$, we can find disjoint open sets E_n such that $\varphi_n \in E_n$, and functions $g_n \in \mathfrak{A}$ with $\operatorname{car}(g_n) \subseteq E_n$. Arrange these functions in an infinite matrix by defining, for example,

$$h_{ij} = g_m, \text{ where } m = 2^{i-1}(2j-1), \qquad i, j = 1, 2, \cdots.$$

Let

$$f_j = \sum_{i=1}^{\infty} \alpha_{ij} h_{ij}, \qquad j = 1, 2, \cdots,$$

where the constants α_{ij} are chosen to make the series converge in $\mathfrak{A}$. Then $f_j f_k = 0$, $j \neq k$, and $f_j \in \overline{\mathfrak{J}(\omega_0)} \sim \mathfrak{J}(\omega_0)$. It follows from Lemma 5.2 that $\nu(f_j^3) = 0$ for all sufficiently large j.

6. An example. We conclude with a discussion of an example, due to C. Feldman [2], which will show that the main boundedness theorem of Section 2 can not be improved.

Let $\mathfrak{A}$ be the commutative Banach algebra which is the completion of the algebra $\mathfrak{A}_0$ of all finite sums

$$\sum_{i=1}^{n} \alpha_i e_i + \gamma r,$$

where α_i and γ are complex, e_i are mutually orthogonal idempotents, $r^2 = 0$, $e_i r = r e_i = 0$, and

$$\| \sum \alpha_i e_i + \gamma r \| = \max\{ [\sum | \alpha_i |^2]^{\frac{1}{2}}, | \gamma - \sum \alpha_i | \}.$$

We refer to [2] for a proof that $\mathfrak{A}$ is a Banach algebra. Let $\mathfrak{R}$ be the one dimensional ideal generated by r. Then $\mathfrak{A}_0/\mathfrak{R}$ is isometrically isomorphic with the algebra of finite sequences $[\alpha_i]$ with the norm $[\sum | \alpha_i |^2]^{\frac{1}{2}}$, so $\mathfrak{A}/\mathfrak{R}$ is the algebra l_2 and $\mathfrak{R}$ is the radical of $\mathfrak{A}$. Feldman proved that there is no closed subalgebra of $\mathfrak{A}$ isomorphic to $\mathfrak{A}/\mathfrak{R}$. Thus a natural generalization of the Wedderburn principal theorem for finite dimensional algebras is false for $\mathfrak{A}$. We shall show, however, that there does exist a non-closed subalgebra $\mathfrak{B}$ of $\mathfrak{A}$ such that $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$. The map ν: $l_2 \to \mathfrak{B}$ will be the discontinuous isomorphism of l_2 which we seek. We summarize the information we need in the following theorem.

6.1. THEOREM. (a) *The span of the idempotents is dense in* $\mathfrak{A}$.

(b) *There exists no closed subalgebra* $\mathfrak{B}'$ *of* $\mathfrak{A}$ *such that* $\mathfrak{A} = \mathfrak{B}' \oplus \mathfrak{R}$.

(c) *There exists a non-closed subalgebra* $\mathfrak{B}$ *of* $\mathfrak{A}$ *such that* $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$.

Proof. Let I denote the algebraic span of the idempotents e_i of $\mathfrak{A}$. To prove (a) it is enough to show r is in $\bar{I}$. Since the topology of l_1 is stronger than that of l_2 one may find sequences $\xi_n = [\alpha_i^{(n)}]$ of non-negative numbers, each having only finitely many non-zero terms, such that the norm of ξ_n approaches one in l_1 and zero in l_2. Let $x_n = \sum \alpha_i^{(n)} e_i$. Then $x_n \in I$ and

$$\| r - x_n \| = \max\{ [\sum | \alpha_i^{(n)} |^2]^{\frac{1}{2}}, | 1 - \sum \alpha_i^{(n)} | \} \to 0. \qquad \text{Thus } \mathfrak{A} = \bar{I}.$$

To prove (b) suppose there exists a closed subalgebra $\mathfrak{B}'$ such that $\mathfrak{A} = \mathfrak{B}' \oplus \mathfrak{R}$. Then $e_i \in \mathfrak{B}'$ for each i, for writing $e_i = b + \gamma r$, $b \in \mathfrak{B}'$, we see $e_i = e_i^2 = (b + \gamma r)^2 = b^2 = b + \gamma r$. Thus $b^2 = b = e_i$, so $\mathfrak{B}' \supseteq I$. Since $\mathfrak{B}'$ is assumed closed, we have $\mathfrak{B}' = \mathfrak{A}$ by (a).

Now $l_1 \subseteq l_2 = \mathfrak{A}/\mathfrak{R}$ and, using Zorn's lemma, we may construct a vector subspace V of l_2 such that $l_2 = l_1 \oplus V$. We construct an isomorphism ν of l_2 into $\mathfrak{A}$ as follows: For $\xi = [\alpha_i] \in l_1$ define

$$\nu(\xi) = \sum \alpha_i e_i.$$

Since $\| e_i \| = 1$ in $\mathfrak{A}$ the series converges absolutely. For $\xi = [\alpha_i] \in V$ we have $\sum | \alpha_i | = +\infty$. Let

$$x_n = \sum_{i=1}^{n} \alpha_i e_i + \sum_{i=1}^{n} \alpha_i r.$$

Then $\{x_n\}$ is a Cauchy sequence in $\mathfrak{A}$ since

$$\| x_m - x_n \| = [\sum_{n+1}^{m} | \alpha_i |^2]^{\frac{1}{2}} \to 0.$$

Define $\nu(\xi) = \lim x_n$ in $\mathfrak{A}$. One shows easily that $(\rho\nu)(\xi) = \xi$, $\xi \in l_2$, where ρ is the Gelfand homomorphism. Let $\mathfrak{C}$ and $\mathfrak{D}$ denote the range of ν on l_1 and V respectively. Then $\mathfrak{C}$ is clearly an algebra. If $\mathfrak{B} = \mathfrak{C} \oplus \mathfrak{D}$ is an algebra, it follows that ν is an isomorphism and $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$. If $x, y \in \mathfrak{D}$ and $x = \lim x_n$, $y = \lim y_n$, where

$$x_n = \sum_{i=1}^{n} \alpha_i e_i + \sum_{i=1}^{n} \alpha_i r, \qquad y_n = \sum_{i=1}^{n} \beta_i e_i + \sum_{i=1}^{n} \beta_i r,$$

then $x_n y_n = \sum_{i=1}^{n} \alpha_i \beta_i e_i$ and $[\alpha_i \beta_i] \in l_1$. Thus $xy \in \mathfrak{C}$. Similarly, if $x \in \mathfrak{D}$, $y \in \mathfrak{C}$, then $xy \in \mathfrak{C}$.

6.2. COROLLARY. *Let $\nu: l_2 \to \mathfrak{B}$ be the discontinuous isomorphism of the last theorem. If p_n is the idempotent $[1, 1, \cdots, 1, 0, 0, \cdots]$ in l_2 whose first n entries are ones, then $\| p_n \| = n^{\frac{1}{2}}$ in l_2, while $\| \nu(p_n) \| = n$ in $\mathfrak{A}$.*

The algebra $\mathfrak{A}$ has an additional interesting property.

6.3. COROLLARY. *The algebra $\mathfrak{A}$ admits two inequivalent complete multiplicative norms.*

Proof. Besides the given norm in $\mathfrak{A}$ we have the norm

$$||| x ||| = || y ||_2 + | \lambda |,$$

where $x = y + \lambda r$, $y \in \mathfrak{B}$, and $\| y \|_2$ is the norm of $\nu^{-1}(y)$ in l_2. The two norms are inequivalent since $\mathfrak{B}$ is not closed in $\mathfrak{A}$.

Added in proof. In Section 4 it was shown that when $\mathfrak{A} = C(\Omega)$, then $\nu = \mu + \lambda$, where μ is a continuous homomorphism and λ maps into the radical of the image algebra. One easily sees that this splitting of ν holds for an algebra $\mathfrak{A}$ whenever the following two conditions hold: (1) There are no closed non maximal primary ideals, i. e. $\mathfrak{M}(\phi) = \overline{\mathfrak{J}(\phi)}$, $\phi \in \Phi_{\mathfrak{A}}$. (2) There is a constant K such that if $\phi \in \Phi_{\mathfrak{A}}$, $g \in \mathfrak{J}(\phi)$, then there exists $h \in \mathfrak{J}(\phi)$ with $gh = g$, $\| h \| \leqq K$. Y. Katznelson has pointed out to us that the algebra $\mathfrak{F}$ of absolutely convergent Fourier series has these properties (cf. N. Wiener,

The Fourier integral and certain of its applications, Cambridge, 1933, page 88). Statements (c) and (d) of Theorem 4.3 carry over. However $\mu(\mathfrak{F})$ will contain radical elements whenever the ideal $\{x \mid \mu(x)=0\}$ is not the kernel of its hull. The question of whether $\mathfrak{F}$ has discontinuous homomorphisms is thus reduced to the open question of the existence of a homomorphism of $\mathfrak{F}$ into a radical Banach algebra. Any such homomorphism is necessarily discontinuous.

One might think that condition (2) would imply (1). However Katznelson has constructed the following elegant example. Let $\mathfrak{A}$ be the algebra of all complex sequences $x=[\xi_0,\xi_1,\cdots]$ such that $\xi_n \to 0$ and $\sup_n n^{-\frac{1}{2}} \sum_{i=1}^{n} |\xi_i - \xi_{i-1}| < \infty$. The norm of x is the sum of this supremum and $\sup_n |\xi_n|$. Adjoin a unit obtaining $\mathfrak{A}_1$. Then $\Phi_{\mathfrak{A}_1}$ is the integers with point at infinity and $\{x \mid \lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{i=1}^{n} |\xi_i - \xi_{i-1}| = 0\}$ is a closed primary ideal. But the idempotents $k_{[0,n]}$ form a bounded system of relative units.

REFERENCES.

[1] P. J. Cohen, "Factorization in group algebras," *Duke Mathematical Journal*, vol. 26 (1959), pp. 199-206.

[2] C. Feldman, "The Wedderburn principal theorem in Banach algebras," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 771-777.

[3] I. Gelfand, "Normierte Ringe," *Matematičeskiĭ Sbornik*, Novaya Seriya, vol. 9 (1941), pp. 3-24.

[4] ——— and M. Naimark, "On the imbedding of normed rings into the ring of operators in Hilbert space," *ibid.*, vol. 12 (1943), pp. 197-213.

[5] I. Kaplansky, "Normed algebras," *Duke Mathematical Journal*, vol. 16 (1949), pp. 399-418.

[6] H. A. Kober, "A theorem on Banach spaces," *Compositio Mathematica*, vol. 7 (1939), pp. 135-140.

[7] L. H. Loomis, *An introduction to abstract harmonic analysis*, New York, 1953.

[8] C. E. Rickart, "The uniqueness of norm problem in Banach algebras," *Annals of Mathematics*, vol. 51 (1950), pp. 615-628.

[9] G. Silov, "On regular normed rings," *Trudy Matematičeskogo Instituta imeni V. A. Steklova*, vol. 21, Moscow, 1947 (Russian).,

[10] M. H. Stone, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

[11] B. Yood, "Topological properties of homomorphisms between Banach algebras," *American Journal of Mathematics*, vol. 76 (1954), pp. 155-167.

ON HYPERSURFACES WITH NO NEGATIVE SECTIONAL CURVATURES.*

By RICHARD SACKSTEDER.[1]

1. Introduction. If a hypersurface $S \subset E^{n+1}$ is the boundary of a sufficiently smooth convex body, S has the intrinsic properties that all of its sectional curvatures are non-negative and as a Riemannian manifold it is complete in the sense of Hopf and Rinow. Most of this paper is concerned with the converse question, that is, given a complete Riemannian manifold with no negative sectional curvatures immersed in E^{n+1}, when is the image the boundary of a convex body?

The second appendix is independent of the body of the paper. In it, some counterexamples are given which show that theorems of Hilbert and Weyl on the extremes of the curvatures of a surface are false without suitable smoothness assumptions.

2. Preliminaries. A Riemannian manifold is said to be of *class* C^k $(k \geqq 1)$ if it is of class C^k as a differentiable manifold and in any coordinate system the components of the metric tensor are functions of class C^{k-1}. Unless the contrary is stated, a manifold will mean a connected manifold without a boundary. Let M_1 and M_2 be manifolds of class C^k and of dimensions m_1 and m_2 $(m_1 < m_2)$ respectively. M_1 will be said to be C^m-*immersed* $(m \leqq k)$ in M_2 if there is a single-valued map $X: M_1 \to M_2$ of class C^m with Jacobian of rank m_1 at every point of M_1. Such an immersion will be called *isometric* if M_1 and M_2 are Riemannian manifolds and the metric induced on

* Received November 25, 1959; revised May 2, 1960.

[1] This paper is an extension of a section of the author's doctoral dissertation written at the Johns Hopkins University under the direction of Professor Philip Hartman. The author gratefully acknowledges Professor Hartman's advice and assistance.

The research at Johns Hopkins was supported by the United States Steel Foundation and by the United States Air Force through the Air Force Office of Scientific Research under Contract No. AF 18(603)-41. The results were extended while the author held an Office of Naval Research postdoctoral fellowship at Yale University.

the image $X(M_1)$ as a subset of M_2 is the same as the metric induced locally from the metric of M_1. If the map F is one-to-one, M_1 will be said to be C^m-*imbedded* in M_2. In the special case in which M_1 is of dimension m and $M_2 = E^{m+1}$, the image $X(M_1)$ under an immersion will be called an *m-hypersurface*, of if $m = 2$, a *surface*.

A Riemannian manifold becomes a metric space if the distance between two points is defined to be the greatest lower bound of lengths of the arcs connecting them. A manifold[2] is said to be *complete* if the metric space obtained in this way is complete. A number of equivalent definitions of completeness will be used below, cf. [11] or [16]. The image of a complete n-manifold isometrically immersed in E^{n+1} is called a *complete n-hypersurface.*

The main result of this paper is the following theorem which generalizes a classical theorem of Hadamard [7].

THEOREM (*). *Let M be a complete Riemannian n-manifold* $(n \geqq 2)$ *and $X: M \to E^{n+1}$ a C^{n+1} isometric immersion of M into E^{n+1}. Suppose that every sectional curvature of M is non-negative and that at least one is positive. Then the image $X(M)$ is the boundary of a convex body in E^{n+1}.*

For sufficiently smooth hypersurfaces this theorem generalizes results of Hadamard [7], Stoker [22], Van Heijenoort [23], and Chern and Lashof [4]. Hadamard's result corresponds to the case where M is compact and all sectional curvatures of M are positive. Stoker generalized Hadamard's result for $n = 2$ by removing the restriction that M is compact. Chern and Lashof obtained another generalization for $n = 2$ in which M is assumed to be compact but the sectional curvatures are only required to be non-negative rather than strictly positive. Finally, Van Heijenoort considered all $n \geqq 2$, but replaced the assumptions on sectional curvatures by local convexity conditions.

The condition that sectional curvatures of M be non-negative is easily seen to be equivalent to the condition that the second fundamental form of $X(M)$ be semi-definite; cf. Appendix 1. However, the non-negativeness of the sectional curvatures does not imply that the second fundamental form remains either non-negative or non-positive at all points and consequently does not imply any sort of local convexity property. This assertion is clear for hypercylinders. A less trivial example is given by the surface in the (x, y, z) space E^3 defined by $z = x^3(1 + y^2)$ for $|y| < \frac{1}{2}$. The second fundamental form of this surface is positive definite for $x > 0$ and negative definite for $x < 0$. It

[2] Hopf and Rinow [11] require that the manifold satisfy the second axiom of countability, but this condition is automatically fulfilled in a connected Riemannian manifold; cf. [20], p. 23.

follows from Theorem (*) that no neighborhood of the origin on this surface can be a part of a complete surface with non-negative Gaussian curvature.

In Theorem (*), the purpose of the assumption that at least one sectional curvature is positive is to eliminate the possibility that $X(M)$ is a non-convex hypercylinder. Consequently, Theorem (*) complements a result of Hartman and Nirenberg [8], p. 912 who proved that if all sectional curvatures of M are zero, $X(M)$ is a hypercylinder.

Let r denote the maximum rank of the second fundamental form of $X(M)$. In Appendix 1, it is shown that, under the hypotheses of Theorem (*), r is determined intrinsically. Certain supplementary facts about $X(M)$ which follow from Appendix 1 and the proof of Theorem (*) are formulated below.

SUPPLEMENT TO THEOREM (*). *Under the hypotheses of Theorem* (*), E^{n+1} *can be decomposed as a product* $E^{n+1}=E^{r+1}\times E^{n-r}$ *in such a way that* $X(M)=P_1X(M)\times P_2X(M)$, *where* P_1 *and* P_2 *denote respectively the orthogonal projections onto* E^{r+1} *and* E^{n-r}. *Then* $P_2X(M)=E^{n-r}$ *and* $P_1X(M)$ *is a hypersurface in* E^{r+1} *which bounds a convex body which contains no complete line. The integer* r *is determined intrinsically and satisfies* $2\leqq r\leqq n$.

The proof of Theorem (*) depends on a series of lemmas and theorems which are proved in the next two sections. The lemmas of Section 3 are general propositions, some of which are known or are implicit in the literature. No attempt is made to state these lemmas in the most general form in which they are valid, because to do so would make the applications of them less clear. The lemmas of Section 4 are more specifically related to the problem of this paper. Most of these lemmas are not of interest in themselves because the truth of their assertions is clear once Theorem (*) has been proved. Two propositions which are perhaps of some interest are formulated as Theorems 1 and 2.

3. Lemmas. Throughout this section M will denote a Riemannian manifold of class C^1 and $X\colon M\to E^{n+1}$ a C^1 immersion of M into E^{n+1}.

The boundary of a set S will be denoted by S'.

Suppose that V is an $(n+1)$-vector. The following definitions and notation will often be used in this and the next section. W will denote an open connected subset of M such that the inner product $N(p)\cdot V\neq 0$ at any point p of W. Here $N\colon M\to S^{n+1}$ is the "composition" of X and the normal map of $X(M)$. It may not be possible to define N in the large on M, but $N(p)\cdot V\neq 0$ has an obvious meaning. Let Π_V denote the hyperplane through the origin orthogonal to V. Let $(x^0,\cdots,x^n)$ be coordinates in E^{n+1} such that

V is a unit vector along the x^0-axis and Π_V is the hyperplane $x^0 = 0$. For p in W, let $X(p) = (z(p), Y(p)) = (x^0(p), \cdots, x^n(p))$ where $z(p)$ is a scalar and $Y: W \to \Pi_V$.

A pair of points p, q of W will be said to *lie on a segment in* W if there is an arc γ in W whose image under Y is the line segment in Π_V joining $Y(p)$ to $Y(q)$. Since Y is a local homeomorphism, γ is unique as a set, and γ will be called *the segment connecting* p *to* q. A subset U of W will be called *W-convex* if every pair of points in U which lie on a segment in W lie on a segment in U. Every component of a W-convex set is W-convex and the intersection of any family of W-convex sets is W-convex. Note that the notion of W-convexity depends on the vector V as well as on W.

LEMMA 1. *Let U be a closed connected subset of W which is W-convex. Then $Y(U)$ is convex and Y is one-to-one on U.*

Proof. The proof is quite similar to proofs which have been given for a theorem of Tietze, cf. [12], p. 448, or [2], p. 56, ex. 22. Consequently it will only be sketched.

The proof depends on the following facts: 1) if p and q are points of U, there are points $p_0 = p, p_1, \cdots, p_k = q$ of U such that p_{j-1} and p_j lie on a segment in U for $j = 1, 2, \cdots, k$, 2) If p_0 and p_1, and p_1 and p_2 lie in a segment in U, then p_0 and p_2 lie on a segment in U. The proofs of 1) and 2) can be carried out as in [2] or [12] in spite of the fact that Y is only locally one-to-one on W. That Y is one-to-one on U follows from the convexity of $Y(U)$, the connectedness of U, and the fact that Y is a local homeomorphism. This completes the proof of Lemma 1.

Let U be a subset of W. By the *W-convex hull of U* is meant the smallest W-convex set containing U. Let $U = U^0$ and define $U^1, U^2, \cdots$ inductively by $U^k =$ the union of all segments in W connecting pairs of points of U^{k-1}. Clearly $U^0 \subset U^1 \subset U^2 \cdots$.

LEMMA 2. *The W-convex hull of U is* $U^\infty \equiv \bigcup_{k=1}^{\infty} U^k$.

The proof of Lemma 2 is quite simple and will be omitted.

LEMMA 3. *Let M be a simply connected manifold and let A and B be closed subsets of M. Then if two points x, y are connected in $M - A$ and in $M - B$, they are connected in $M - (A \cup B)$.*

Lemma 3 is a special case of a known theorem, cf. [25], p. 242, Theorem 9.2. It can also be verified directly using the fact that it holds for $M = E^2$.

COROLLARY 1. *Let M be a simply connected manifold and let D be an open connected subset. Then each component of $M - D$ contains just one component of the boundary of D.*

COROLLARY 2. *Let M be a simply connected manifold and D an open subset of M. Suppose that C_1 and C_2 are disjoint components of D. Then there is a component B of $M - D$ which separates C_1 and C_2 in M and which contains a component of the boundary of C_1 which also separates C_1 and C_2 in M.*

Corollaries 2 and 2 follow from Lemma 3 by exactly the arguments used in [17] to prove Theorems 14.5 and 14.3, respectively; cf. [25], p. 47 ff.

LEMMA 4. *Let M_1 be an open subset of a simply connected manifold M. Suppose that C is a component of M_1. Let C_0 be the component of $C \cup (M - M_1)$ containing C and let C^* be the union of C with all of the components of $M - M_1$ which intersect the closure of C. Then $C_0 = C^*$.*

Proof. Clearly $C_0 \supset C^*$. Let $\{U_a\}$ be the components of $M - C$. Then, if each of the sets $U_a \cap C_0$ is connected, $C_0 = C^*$. On the other hand if, for example, $U_1 \cap C_0$ is not connected, $U_1 \cap C_0 = A \cup B$ where A and B are closed, disjoint and non-void. Since U_a' is connected by Corollary 1, it can be supposed that $U_1' \subset A$. Then $C_0 = (C_0 - B) \cup B$ and $C_0 - B$, B are closed, disjoint, and non-void. This contradicts the connectedness of C_0 and completes the proof.

LEMMA 5. *Let M be a complete Riemannian n-manifold and $X: M \to E^{n+1}$ an isometric immersion of class C^1. Suppose that $B_1, B_2, \cdots$ is a sequence of subsets of M which have the properties:* (a) *$X(B_i)$ is convex,* (b) *$X \mid B_i$ is a homeomorphism,* (c) *$\limsup B_i$ is non-empty, say $p \in \limsup B_i$,* (d) *$L = \lim X(B_i)$ exists. Then there is a subset B of $\limsup B_i$ containing p which is such that $X \mid B$ is a homeomorphism and $X(B) = L$.*

Note: Here $\limsup B_i$ and $\lim X(B_i)$ are used in the sense of [26], p. 10.

Proof. By (c), it can be supposed that $B_1, B_2, \cdots$ is such that there is a sequence of points $p_1, p_2, \cdots$ satisfying $p_i \in B_i$ and $\lim p_i = p$ as $i \to \infty$. If $L \neq X(p)$, let $q \neq X(p)$ be a point of L, and let $q_1, q_2, \cdots$ be a sequence of points such that $q_i \in B_i$ and $q = \lim X(q_i)$ as $i \to \infty$. L is clearly convex, in particular, the segment $X(p)q$ connecting $X(p)$ to q is in L. Let coordinates $(x^0, x^1, \cdots, x^n)$ be chosen in E^{n+1} such that $X(p) = (0, \cdots, 0)$ and $q = (1, 0, \cdots, 0)$. Let $p_t = (t, 0, \cdots, 0)$. Then p_t is in L for $0 \leqq t \leqq 1$.

Let s denote the largest t value such that the segment p_0p_s, open at p_s, has a homeomorphic preimage containing p.

First it will be shown that $s > 0$. Let U be a compact neighborhood of p such that $X \mid U$ is a homeomorphism and let C_i be the component of $U \cap X^{-1}(p_iq_i)$ containing p_i. Then for large i, $X(C_i)$ is a non-degenerate closed segment. Let r_i be the point of C_i which maps into the endpoint of $X(C_i)$ which is not $X(p_i)$. Since U is a compact neighborhood of p it can be assumed that $\lim r_i = r$ exists as $i \to \infty$ and $r \neq p$. Since $X \mid U$ is a homeomorphism and $\lim X(C_i) = X(p)X(r)$ exists, it follows that $\lim C_i = C$ exists and $X(C)$ is the segment $X(p)X(r)$. Clearly this segment will be a subset of $X(p)q$. This proves that $s > 0$.

Now it will be proved that $s = 1$, i.e., that $p_s = q$. If $s \leqq 1$, the half open segment p_0p_s has a homeomorphic preimage in M. Since M is isometric and M is complete the closed segment p_0p_s has a homeomorphic preimage in M containing p. But now the argument just used to prove that $s > 0$ can be applied again to show that if $s < 1$, s can be increased. Consequently, $s = 1$, and $X(p)q$ has a homeomorphic preimage in M which contains p. The preimage is easily seen to be unique.

Let B be the union of all of the preimages corresponding to all of the points q in L. Then $X(B) = L$ and since B is connected and X is a local homeomorphism, $X \mid B$ is a homeomorphism. This completes the proof of Lemma 5.

4. Flat points on *M*.

Let (H) denote the hypothesis: (i) M is a complete Riemannian n-manifold $(n \geqq 2)$ of class C^{n+1}, (ii) $X\colon M \to E^{n+1}$ is an isometric immersion of class C^{n+1}, (iii) all of the sectional curvatures of $X(M)$ are non-negative.

Let $h_{ij}(p)$ be the coefficients of the second fundamental form of $X(M)$ at the point p of M, relative to some fixed local coordinate system and choice of the unit normal vector $N\colon M \to S^n$. Let M_0, M_1 be the subsets of M defined as follows:

$$M_0 = \{p : p \in M, h_{ij}(p) = 0, 1 \leqq i, j \leqq n\}, \quad M_1 = M - M_0.$$

Points in M_0 will be called *flat points.*

LEMMA 6. *Assume (H). Let T be a connected subset of M_0. Then the normal N is constant on T and $X(T)$ lies in a hyperplane orthogonal to N on T.*

Proof. At a flat point of M the rank of the map $N: M \to S^n$ is zero. Since N is a map of class C^n a theorem of Sard [21], p. 888, implies that $N(M_0)$ is a one-dimensional zero set, in particular, $N(M_0)$ is totally disconnected. This proves that $N(T)$ consists of a single point, N_0. Let $(u^1, \cdots, u^n)$ be local coordinates near a point p_0 of T and let $f(u^1, \cdots, u^n) = N_0 \cdot (X(p) - X(p_0))$. Then $\partial f / \partial u^i = 0$ for $i = 1, \cdots, n$ on T, hence, by a theorem of A. P. Morse [15], f is constant on T. It follows that $N_0 \cdot X(p) = N_0 \cdot X(p_0)$ on T, that is $X(T)$ lies in the hyperplane $N_0 \cdot (x^1, \cdots, x^n) = N_0 \cdot X(p_0)$. This proves Lemma 6.

Remark. It is only for the purpose of proving Lemma 6 that the immersion in Theorem (*) is required to be of class C^{n+1}, instead of C^2. The non-negativeness of the sectional curvatures of M was not used in the proof of Lemma 6.

LEMMA 7. *Assume (H) and suppose that M is simply connected. Let T be a component of M_1. Let $W \subset M$ be as in Section 3. Then the intersection of each component of $M - T$ with W is W-convex.*

Proof. Let $\{U_a\}$ be the compents of $M - T$. Then by Corollary 1, U_a contains only one component of $T' = (M - T)'$, hence U_a' is connected. Lemma 6 implies that the normal to $X(M)$ is constant on U_a', say $N \equiv N_a$ on U_a' and there is a hyperplane Π_a with normal N_a such that $X(U_a') \subset \Pi_a$.

As in Section 3, let V be an $(n+1)$-vector such that for p in W, $N(p) \cdot V \neq 0$, let Π be a hyperplane in E^{n+1} orthogonal to V, and suppose that E^{n+1} has orthogonal coordinates $(x^0, \cdots, x^n)$ such that Π is $x^0 = 0$. If $N_a \cdot V \neq 0$, let $z(x^1, \cdots, x^n; a)$ denote the x^0 coordinate of the point on Π_a whose orthogonal projection onto Π is $(0, x^1, \cdots, x^n)$. Define the function $g: W \to E^1$ by $g(p) = x^0(p)$ if p is in $W \cap T$, $g(p) = z(Y(p); a)$ if p is in $W \cap U_a$, where $X(p) = (x^0(p), \cdots, x^n(p))$ and $Y(p) = (x^1(p), \cdots, x^n(p))$. Then g is of class C^2 by Lemma 6, since $U_a' \subset M_0$.

It can be supposed that the second fundamental form of $X(M)$ is non-negative semi-definite in T. Let a be fixed, let $J(p; a) = g(p) - z(Y(p); a)$, and define a (not necessarily isometric) immersion $G_a: W \to E^{n+1}$ by $G_a(p) = (J(p; a), Y(p))$. Then G_a is a C^2 immersion, the flat points of the immersion are precisely the points of $W - T$, and the second fundamental form of $G_a(W)$ is non-negative semi-definite. Let V_a be the W-convex hull of $U_a \cap W$. To prove that $U_a \cap W$ is W-convex it suffices to show that all points of V_a are flat points of $G_a(W)$, so that V_a does not meet T and, hence, $V_a = V_a \cap W$.

Let $U^0 = U_a \cap W$ and define $U^1, U^2, \cdots, U^\infty$ as in Lemma 2. Then

$V_a = U^\infty$ and to show that V_a consists entirely of flat points it suffices to prove that $U^0 = U^1$, hence $U^0 = U^\infty$. First it will be proved that

$$(1) \qquad J(p;a) \equiv 0 \text{ on } U^1.$$

Note that by definition of $J(p;a)$, $J(p;a) \equiv 0$ on U^0. If p_1, p_0 are points of U^0 which lie on a segment in W, let p_t denote the point on the segment which is such that $Y(p_t) = tY(p_1) + (1-t)Y(p_0)$, and let $H(t) = J(p_t;a)$ for $0 \leqq t \leqq 1$. $H''(t) \geqq 0$ because the second fundamental form of $G_a(W)$ is non-negative. $H(0) = H(1) = 0$ and $H'(0) = 0$ because p_1 and p_0 are in U^0. These statements imply $H(t) \equiv 0$ for $0 \leqq t \leqq 1$. This proves (1).

To complete the proof that $U^0 = U^1$, note that Y is one-to-one in a neighborhood S of the segment connecting p_1 to p_0. Then there is a function f defined in $Y(S)$ such that $J(p;a) \equiv f(Y(p))$ for p in S. It can be assumed that $Y(S)$ is convex. Then f will be a convex function. Let Δ^j denote the n-vector whose j-th component is Δ and other components are 0. Fix t, $0 < t < 1$ and put

$$Q(\Delta) = tf(Y(p_1) + \Delta^j) + (1-t)f(Y(p_0) + \Delta^j) - f(Y(p_t) + \Delta^j).$$

Q is defined for small Δ and by the convexity of f, $Q \geqq 0$. By (1), $Q(0) = 0$, so Q has a minimum at $\Delta = 0$. Therefore

$$Q''(0) = tf_{jj}(Y(p_1)) + (1-t)f_{jj}(Y(p_0)) - f_{jj}(Y(p_t)) \geqq 0.$$

Since p_1 and p_0 are in U^0, $f_{jj}(Y(p_1)) = f_{jj}(Y(p_0)) = 0$ and so $f_{jj}(Y(p_t)) \leqq 0$. On the other hand, $f_{jj}(Y(p_t)) \geqq 0$ because f is convex. Therefore $f_{jj}(Y(p_t)) = 0$ for $j = 1, \cdots, n$ and p_t is a flat point of $G_a(W)$, hence p_t is not in T and $U^0 = U^1$. This shows that V_a consists entirely of flat points of $G_a(W)$ and completes the proof of Lemma 7.

THEOREM 1. *Assume* (H). *Let C be a component of the set of flat points of M. Then $X(C)$ is convex and $X \mid C$ is a homeomorphism.*

Proof. Without loss of generality, M can be assumed to be simply connected, for otherwise M can be replaced by its universal covering manifold. By Lemma 6, N is constant on C. It can be supposed that $N \equiv V$ on C. Let W be the component of the set $\{p: p \in M, N(p) \cdot V \neq 0\}$ containing C. Let $\{T_a\}$ be the components of M_1 and let U_a be the component of $M - T_a$ containing C. Lemma 7 shows that $W \cap U_a$ is W-convex; hence the component of the intersection of all of these sets which contains C is W-convex. This component is a connected subset of M_0 hence it is C. This proves that C is W-convex. Lemma 1 shows that $Y(C)$ is convex, where $Y\colon W \to \Pi_V$ is as in

Section 3. V is the normal to $X(M)$ on C, hence $Y(C)$ is just a translation of $X(C)$. This proves that $X(C)$ is convex and since C is connected, $X \mid C$ is a homeomorphism. This completes the proof of Theorem 1.

LEMMA 8. *Assume (H) and suppose that M is simply connected. Let C_1 and C_2 be distinct components of M_1. Then there is a subset L of C_1' such that L separates C_1 from C_2 in M and $X(L)$ is an $(n-1)$-flat.*

Proof. By Corollary 2, there is a component B of M_0 which separates C_1 from C_2 and which contains a component L of C_1' which also separates C_1 from C_2. B is isometric to a convex subset of E^{n+1} by Theorem 1. By considering all possible types of convex subsets of dimension n or less (cf. [3], p. 3), it is easy to verify that B can separate a complete n-manifold only if B is isometric to an $(n-1)$-flat or to a set bounded by two parallel $(n-1)$-flats in E^n. (A set of the latter type will be called a *slab* below.) L is a connected subset of B' which separates C_1 from C_2. Clearly $X(L)$ must be an $(n-1)$-flat in either case.

LEMMA 9. *Assume (H) and suppose that M is simply connected. Suppose that C is a component of M_1 and that C_0 is the component of $M_0 \cup C$ containing C. Let p be a point of C_0'. Then there is a subset L of C' lying in the same component of M_0 as p which is such that $X(L)$ is an $(n-1)$-flat.*

Proof. Let B be the component of M_0 containing p. Let $p_1, p_2, \cdots$ be a sequence of points in $M_1 - C$ such that $p = \lim p_i$ as $i \to \infty$. By Lemma 8 there is a component B_i of C' which is such that $X(B_i)$ is an $(n-1)$-flat and B_i separates p_i from C. Either $B_j \subset B$ for some j or $B_i \cap B$ is empty for all i. In the first case, let $L = B_j$ and the lemma is proved for this case.

If $B_i \cap B$ is empty for all i, it will first be shown that $p \in \limsup B_i$. To see this, let q be a point of $B \cap C'$ (cf. Lemma 4). Let pq be the inverse image in B of the segment $X(p)X(q)$. Then B_i does not intersect pq for any i, hence B_i separates p from p_i. A simple application of Lemma 5 shows that there is a set $L \subset \limsup B_i \subset C'$ such that $X(L)$ is an $(n-1)$-flat. This completes the proof of Lemma 9.

LEMMA 10. *Assume the conditions of Lemma 9. Then there is a subset K of C_0 containing p such that $X(K)$ is an $(n-1)$-flat. In addition there is a neighborhood U of p with the properties:* (i) *$U - K$ has exactly two components D_1 and D_2,* (ii) *D_1 and D_2 are homeomorphic to n-cells.* (iii) *$D_1 \subset C_0$,* (iv) *$D_2 \cap C_0 = 0$. Finally, if Π is any hyperplane which is not parallel to $N(p)$, U can be chosen such that the orthogonal projection of*

$X(U \cap C_0)$ onto Π is a solid n-hemisphere (including the equatorial hyperplane).

Proof. Let B be the component of M_0 containing p. It follows from Theorem 1 and Lemma 9 that $X(B)$ is either an $(n-1)$-flat or a slab. In either case there is a subset $K \subset B \subset C_0$ such that p is in K and $X(K)$ is an $(n-1)$-flat. Let U be a neighborhood of p which is so small that $X \mid U$ is a homeomorphism and $X(U)$ has a homeomorphic orthogonal projection into a hyperplane $\Pi \subset E^{n+1}$. Let $Y: U \to \Pi$ denote the composition of X with the projection. It can be supposed that $Y(U)$ is an open convex subset of Π and in case $X(B)$ is a slab that U intersects only one component of B'.

Clearly U satisfies the conditions (i) and (ii) above. It remains to show that U can be chosen to satify (iii) and (iv). (The final assertion will be clear from the proof.) Consider the two cases $p \notin C'$ and $p \in C'$. In the first case, Lemma 9 shows that $X(B)$ is a slab. Then one of the components of $U - K$ will lie in $B \subset C_0$ because U only intersects one component of B'. Denote this component by D_1 and the other by D. We have shown that $D_1 \subset C_0$ if p is not in C'. If p is in C', denote by D_1 a component of $U - K$ which contains points of C arbitrarily close to p and let D be the other component. It will be shown that $D_1 \subset C_0$ in this case also, provided that U is small enough, i.e., it will be shown that there are no points in $D_1 - C_0$ arbitrarily close to p.

Suppose that there is a sequence of points $p_1, p_2, \cdots$ of $D_1 - C_0$ such that $p = \lim p_i$ as $i \to \infty$. It can be assumed that the points p_i are in $D_1 - C_0 - M_0$, because if all points of D_1 near p are in M_0 then all points of D_1 near p are in $B \subset C_0$. Lemma 8 implies that there are subsets $B_i \subset C'$ which separate p_i from C and are such that $X(B_i)$ is an $(n-1)$-flat. Then p_i and $C \cap U$ are in separate components of $U - B_i$. In fact, $U - B_i$ has exactly two components, one of which contains p_i while the other contains C. Also, $D \cup K^*$ is in the component of $U - B_i$ containing C, where $K^* \equiv K \cap U$. To verify this note that B_i intersects the connected set D_1 because B_i separates $p_i \in D_1$ from C and C intersects D_1 by definition of D_1. If B_i intersects D, B_i intersects K^* and $X(B)$ is a slab. Then p is an interior point of B, which contradicts $p \in C'$. Hence B_i does not intersect D which along with $B_i \cap D_1 \neq 0$ implies that B_i does not intersect K^*. Since p is in $C' \cap K$ this shows that $D \cup K^*$ is in the component of $U - B_i$ containing $C \cap U$.

Let F denote the intersection of all of the components of $U - B_i$ which contain $C \cap U$. Then $C \cap U \subset F$ and $D \cup K^* \subset F$. $p = \lim p_i$ implies that $Y(p)$ is a boundary point of the convex set $Y(F) \supset Y(D)$. This shows that

$F = D \cup K^* = U - D_1 \supset U \cap C$ which contradicts the definition of D_1. This proves that U can be chosen such that (iii) is satisfied.

To verify that U can be chosen such that (iv) is satisfied, first note that D cannot contain points of C arbitrarily close to p because then for a suitable choice of U, $D \subset C_0$ by the argument just completed and $U \subset C_0$ which contradicts $p \in C_0'$. Therefore suppose that U is so small that $D \cap C$ is empty. If (iv) does not hold for a suitable choice of U there are points of $D \cap C_0'$ arbitrarily close to p. Let $q \in D \cap C_0'$ and let S denote the interior of the convex hull of $Y(K^*) \cup Y(q)$ and let $D_2 = Y^{-1}(S)$. It will be shown that D_2 contains no points of C_0' and this will complete the proof of Lemma 10.

If there is a point r of $C_0' \cap D_2$, let B_r be the component of M_0 containing r. Lemma 9 shows that there is a subset L_r of $B_r \cap C'$ such that $X(L_r)$ is an $(n-1)$-flat. L_r does not intersect D because $D \cap C'$ is empty. Theorem 1 shows that $X(B_r)$ is a slab, hence $Y(B_r \cap U)$ is the intersection of $Y(U)$ with a closed half space of Π. B_r cannot intersect K, because this would imply that $K \subset B_r$, and that p is an interior point of B_r. This shows that q is in the interior of B_r because otherwise $Y(r)$ could not be in the interior of the convex hull of $Y(K^*) \cup Y(q)$. But if q is in the interior of B_r, q is not in C_0'. This contradiction proves Lemma 10.

LEMMA 11. *Assume (H). Let L be a subset of M such that $X(L)$ is k-flat, $0 < k < n$. Then the normal to $X(M)$ is constant on L.*

Proof. It is sufficient to prove the lemma for the case $k = 1$. Let p be on L and let E^{n+1} have coordinates $(x^0, \cdots, x^n)$ such that $X(p)$ is the origin, the unit normal to $X(M)$ is in the positive x^0-direction and $X(L)$ is the x^1 axis. Then, near p, $X(M)$ can be represented by $x^0 = z(x^1, \cdots, x^n)$. The matrix of second partial derivatives of z will be semi-definite and $z_{11} \equiv 0$ near p on L. But these conditions are easily seen to imply that $z_{1j} \equiv 0$ near p on L for $j = 1, \cdots, n$. This implies that the normal is constant on L near p, hence on all of L. ∎

LEMMA 12. *Assume the conditions of Lemma 9. Let L be a subset of C_0, such that $X(L)$ is a k-flat $0 < k < n$. Then every point p of C_0 belongs to a subset $L_p \subset C_0$ such that $X(L_p)$ is a k-flat parallel to $X(L)$.*

Proof. It suffices to prove the case $k = 1$. Let C^* denote the set of points p which belong to a subset L_p of C_0 such that $X(L_p)$ is a 1-flat parallel to the 1-flat $X(L)$. Lemma 5 implies that C^* is closed. To see that C^* is open in C_0 it is sufficient to show that if $q \in L$, there is a neighborhood V of q such that $V \cap C_0 \subset C^*$.

It can be supposed that E^{n+1} has orthogonal coordinates $(x^0,\cdots,x^n)$ such that $X(q)$ is the origin, $N(q)=e$ is the unit vector in the positive x^0-direction, and $X(L)$ is the x^1-axis. Let Π denote the hyperplane $x^0=0$, let W be the component of $\{p: p\in M, N(p)\cdot e_0\neq 0\}$ containing L, and $W_0=W\cap C_0$. Such a component exists by Lemma 11. Define the map $Y\colon W\to\Pi$ by $Y(p)=(x^1(p),\cdots,x^n(p))$, where $X(p)=(x^0(p),\cdots,x^n(p))$. Let $V\subset W$ be a neighborhood of q such that $Y(V\cap C_0)$ is either a solid n-sphere or an n-hemisphere (including the equatorial hyperplane). Such a neighborhood exists by Lemma 10. Let p^* be an arbitrary point of $V\cap(C_0-L)$. It can be supposed that $Y(p^*)=(0,1,0,\cdots,0)$ and that for some $\epsilon>0$ there is a point $p\in V\cap(C_0-L)$ such that $Y(p)=(0,1+\epsilon,0,\cdots)$. Define q_t as the unique point of L such that $x^1(q_t)=t$ and let $T_t\subset\Pi$ be the solid closed triangle $Y(q)Y(p)Y(q_t)$. Define S_t to be the component of the inverse image $Y^{-1}(T_t)$ containing q, and let s, $0\leqq s\leqq+\infty$, be defined by $s=\sup\{t: Y(S_t\cap W_0)=T_t\}$. If $t<s$, $Y\mid S_t$ is a homeomorphism and if $t'\leqq t$, $S_{t'}\subset S_t$.

Now it will be proved that $s=+\infty$. Clearly $s>0$ because of the form of $V\cap C_0$. Suppose, if possible, that $s<+\infty$. Let $S^*=\cup\{S_t: 0<t<s\}$. First it will be shown that S^* is bounded, i.e. that the intrinsic distance between pairs of points of S^* is uniformly bounded. $X(S^*)$ can be represented in the form $x^0=z(x)$ for $x=(x^1,\cdots,x^n)$ in $Y(S^*)$. z is a convex (or concave) function of x on the interior of $Y(S^*)$. If z is convex, then $0\leqq z(x)\leqq z(Y(p))$ for x in $Y(S^*)$ because $z\equiv 0$ and $\partial z/\partial x^i\equiv 0$ on $Y(L)$. A similar inequality holds if z is concave. In either case, $|z(x)|\leqq|z(Y(p))|$ for x in $Y(S^*)$. The intrinsic distance between two points u_1 and u_2 of S^* satisfies the inequality

$$\begin{aligned}\operatorname{dist}(u_1,u_2)&\leqq\operatorname{dist}(Y(u_1),Y(u_2))+|z(Y(u_1))|+|z(Y(u_2))|\\&\leqq\operatorname{dist}(Y(p),Y(q_s))+2\,|z(Y(p))|.\end{aligned}$$

This proves that S^* is bounded. The completeness of M implies that every infinite subset of S^* has a limit point in M. Let S_0 be the closure of S^*. It is easy to verify that Y can be extended to homeomorphism of S_0 onto the closed triangle, $Y(q)Y(p)Y(q_s)$ using the results just proved.

Denote by pq_s the inverse image in S_0 of the segment $Y(p)Y(q_s)$. If the interior of the pq_s (i.e., the inverse image of the interior of the segment $Y(p)Y(q_s)$) is contained in the intersection of the interior of C_0 and W, then it is easy to verify that $Y(S_t\cap W_0)=T_t$ for some $t>s$. This is impossible by definition of s, hence the interior of pq_s intersects either W' or C_0'.

Let r be the point in the intersection of pq_s with $W' \cup C_0'$ such that the distance from $Y(p)$ to $Y(r)$ is as small as possible. Note that $r \neq p, q_s$.

There are two cases to consider, $r \in W'$ and $r \in W \cap C_0'$. If r is in W', let E^{n+1} have orthogonal coordinates $(y_0, \cdots, y_n)$ such that $N(r)$ is the unit vector in the positive y_0 direction and $X(r)$ is the origin. Note that $N(r)$ is orthogonal to the 2-flat determined by $Y(p)Y(r)$ and the unit vector e_0. To verify this, let u be the orthogonal projection of $N(r)$ onto the 2-flat. Then $N(r) \cdot e_0 = 0$ implies that $u \cdot e_0 = 0$, and it remains to verify $u \cdot Y(p)Y(r) = 0$. The image under X of pq_s is a convex curve in the 2-flat with orthogonal projection $Y(p)Y(q_s)$ on Π: $x^0 = 0$, and u is normal to the curve at $X(r)$. If $u \cdot Y(p)Y(r) \neq 0$, the tangent to the curve is in the e_0 direction at $X(r)$ and the curve has an inflection point at $X(r)$. Since convex curves do not have inflection points this is a contradiction. This proves that the 2-flat is a subset of the hyperplane $y_0 = 0$, in particular, the image of pr under X is in $y_0 = 0$.

By Lemma 10 there is neighborhood U of r such that the orthogonal projection of $X(U \cap C_0)$ onto $y_0 = 0$ is either a solid n-sphere or a hemisphere. Call the projected set R. It can be assumed that $X(U)$ is represented by $y_0 = w(y)$, $y = (y_1, \cdots, y_n)$ for y in R. The function w will be convex (or concave) in R and will satisfy $\dot{w} = \partial w/\partial y^i = 0$ for $1 \leqq i \leqq n$ at $y = 0$, hence w does not change sign in R. Let v be any point of U on pr. Then $y_0(v) = 0$. Such a point v cannot be in the interior of C_0 because $N(v) \neq N(r)$ by definition of r, hence $y_0(v) = 0$ implies that y_0 changes sign near v, and w changes sign near the image of v. On the other hand, if there are no points of pr in the interior of C_0, r is in C_0' and all points u of pr near r are in the set K defined in Lemma 10. By Lemma 11, N is constant on K, hence $N(v) = N(r)$ for v in K. This contradicts the definition of r and proves that $r \in W'$ cannot occur.

Now consider the second possibility, that is, that r is in $W \cap C_0'$. Let U and K be as in the conclusion of Lemma 10. Then there is a subsegment of $Y(p)Y(q_s)$ centered at $Y(r)$ and contained in $Y(K \cap U)$. Since $U \cap K \subset U \cap C_0'$, this implies all points on pr near r are in C_0', which contradicts the definition of r. This completes the proof that $s = +\infty$.

It is clear from the proof that $s = +\infty$ that $\inf\{t\colon Y(S_t \cap W_0) = T_t\} = -\infty$. This shows that $Y(W_0)$ contains a strip

$$-\infty < x^1 < +\infty,\ 0 \leqq x^2 < 1 + \epsilon,\ x^i = 0 \text{ for } i > 2.$$

It can be supposed that z is a convex function on the strip, hence $z \geqq 0$ on the strip. Now it will be proved that $\partial z/\partial x^1 = 0$ along the line $x^2 = 1, x^i = 0$

for $i > 2$. Suppose if possible that $\partial z/\partial x^1 \neq 0$ at $x_t = (t, 1, 0, \cdots)$. Then since $z = 0$ on $Y(L)$, the convexity of z implies that for all v, $-\infty < v < +\infty$ $z(t+v, 0, 0, \cdots) = 0 \geqq z(x_t) + v\partial z/\partial x^1(x_t) + 1 \cdot \partial z/\partial x^2(x_t)$. Letting v approach $\pm\infty$ gives a contradiction. This shows that $\partial z/\partial x^1 = 0$ along the line $x^2 = 1$, $x^i = 0$ for $i > 2$, hence $z(t, 1, 0, \cdots)$ is constant for $-\infty < t < +\infty$. Let L^* denote the component of the inverse image of the line $x^2 = 1$, $x^i = 0$, $i > 2$ under Y which contains p^*. Then $X(L^*)$ is a 1-flat parallel to the x^1 axis. This completes the proof of Lemma 12 because p^* was an arbitrary points of C_0 in a neighborhood of q.

THEOREM 2. *Assume (H). Then either the set M_1 of non-flat points of M is connected or $X(M)$ is a hypercylinder.*

Proof. There is no loss of generality in assuming that M is simply connected, for otherwise M can be replaced by its universal covering manifold. Suppose that M_1 is not connected and let C be any component of M_1. Lemma 8 implies that there is a subset $L \subset C'$ such that $X(L)$ is an $(n-1)$-flat. Lemma 12 implies that every point p of C is in a set L_p such that $X(L_p)$ is a $(n-1)$flat. This shows that the rank of the second fundamental form of $X(M)$ is less than two at p, hence the maximum rank of the second fundamental form is one. Then Theorem III of Hartman and Nirenberg [8] shows that $X(M)$ is a hypercylinder. This proves Theorem 2.

5. Proof of Theorem (*). Let r denote the maximum rank of the second fundamental form of $X(M)$. Then by Appendix 1 and the assumption that M has at least one positive section curvature, r satisfies $2 \leqq r \leqq n$ and so $X(M)$ is not a hypercylinder. Let q be a point of M where the rank of the second fundamental form is r. Then the completeness of M and Lemma 2 of [8] or Lemma 2 of [4] can be used to show that there is a subset $L_q \subset M$ containing q and such that $X(L_q)$ is an $(n-r)$-flat. The argument which is essentially the same as that given at the beginning of the proof of Theorem III in [8], p. 913, will not be repeated here. By Theorem 2 above, M_1 is connected, hence Lemma 12 shows that every point p of M is in a subset $L_p \subset M$ such that $X(L_p)$ is an $(n-r)$-flat parallel to $X(L_q)$.

Let $(x^0, \cdots, x^n)$ be orthogonal coordinates in E^{n+1} such that $X(q)$ is the origin and $X(L_q)$ is the set $x^i = 0$ for $0 \leqq i \leqq r$. Let P_1 and P_2 denote respectively the orthogonal projections which map a point $(x^0, \cdots, x^n)$ of E^{n+1} to $(x^0, \cdots, x^r) \in E^{r+1}$ and $(x^{r+1}, \cdots, x^n) \in E^{n-r}$. At any point p of M the normal vector $N(p)$ is clearly orthogonal to $X(L_p)$. It follows easily that the image $P_1X(M)$ is an r-dimensional manifold M^* immersed in E^{r+1} where

$2 \leqq r \leqq n$. Let $X^*: M^* \to E^{r+1}$ be the immersion map. The connectedness of M_1 implies that M is locally convex under X, hence M^* is locally convex under X^* in the sense of Van Heijenoort [23]. Since the rank of the second fundamental form of $X(M)$ is r at q, M^* is absolutely convex at the point of M^* corresponding to q. The theorem of Van Heijenoort [23], p. 241, implies that $X^*(M^*)$ is the boundary of a convex body B^* in E^{r+1}. It follows that $X(M)$ is the boundary of the convex body $B^* \times E^{n-r}$ in E^{n+1}. This completes the proof of Theorem (*). The assertions of the supplement to Theorem (*) are now clear.

6. A remark on the proof of Theorem (*). In the proof of Theorem (*), it was shown that every point p of M was contained in a subset L_p of M such that the normal is constant on L_p and all of the sets $X(L_p)$ are parallel $(n-r)$-flats, where $2 \leqq r \leqq n$. Hartman and Nirenberg proved the corresponding fact for $0 \leqq r < 2$ in the proof of their Theorem III of [8]. One might suspect that a similar result would hold if the semi-definiteness of the second fundamental form were replaced by assumption that the Jacobian of the spherical image map has a rank $\leqq r$. An example will be given here to show that no such conjecture holds.

Let E^4 have coordinates (x^0, x^1, x^2, x^3) and consider the hypersurface defined by $x^0 = z(x^1, x^2, x^3) = x^1 \sin(x^3) + x^2 \cos(x^3)$. Then it can be verified that the rank of the spherical image map is two at every point of this hypersurface. The normal to the hypersurface is constant only on the lines ($=$ 1-flats) determined by $x^3 =$ const. and $x^1 \cos(x^3) + x^2 \sin(x^3) =$ const., but these lines are not parallel.

7. Applications. (a) *A theorem of Lichnerowicz.* Lichnerowicz [14], p. 221, has proved a theorem on the Betti numbers of an orientable n-manifold imbedded in E^{n+1} with a definite second fundamental form. A much stronger conclusion follows immediately from Theorem (*), provided the imbedding is smooth enough.

(b) *Hypersurfaces with a homeomorphic projection onto a hyperplane.* Suppose that a complete hypersurface in E^{n+1} can be represented by a function $x^0 = z(x^1, \cdots, x^n)$ where z is defined and of class C^{n+1} on the hyperplane $x^0 = 0$. Then if the Hessian matrix $(\partial^2 x / \partial x^i \partial x^j)$ is semi-definite and at at least one point is of rank greater than one, Theorem (*) implies that the hypersurface bounds a convex body. The function z will be convex and nonlinear, hence z is not $o(r)$ as $r \to \infty$, where $r^2 = \sum_{i=1}^{n} (x^i)^2$. If $n = 2$, $x^1 = x$,

$x^2 = y$, this shows that if $z_{xx}z_{yy} - z_{xy}^2 \geqq 0$ and inequality holds at one point, then z is not $o(r)$ as $r \to \infty$. This result complements a theorem of S. Bernstein (cf. E. Hopf [10]) in which the same conclusion is drawn from $z_{xx}z_{yy} - z_{xy}^2 \leqq 0$ with inequality at one point.

(c) *The Rigidity of Surfaces.* First, note that a slightly stronger version of Theorem (*) has actually been proved. For, the assumption that the isometry X is of class C^{n+1} was only used to prove that $X(M)$ is of class C^{n+1}. It would have been sufficient to assume that the isometry is of class C^2 and $X(M)$ is of class C^{n+1} as a differentiable manifold.

In view of this remark, the proof of Theorem (*) has the following corollary.

COROLLARY 3. *Let S_1 be a C^2 n-hypersurface which bounds a convex body in E^{n+1} and is not isometric to E^n. Let S_2 be an n-hypersurface of class C^{n+1} which is a C^2 isometric immersion of S_1 in E^{n+1}. Then S_2 bounds a convex body in E^{n+1}. Then S_2 bounds a convex body in E^{n+1}.*

The statement of Corollary 3 has meaning even if S_2 and the isometry are only continuous. This raises the questions: For which k, $1 < k < n+1$ is Corollary 3 correct if S_2 is of class C^k rather than of class C^{n+1}? For which k, $0 < k < n+1$ is Corollary 3 correct if S_2 is of class C^k and the isometry is of class C^1? The analogous question is false if S_2 and the isometry are only continuous, since, for example, a cap can be cut from a sphere, inverted and replaced. It seems likely that Theorem (*) and Corollary 3 are correct if C^k replaces C^{n+1} for $k \geqq 2$. On the other hand, the possibility that the statements become false for $k = 1$ is suggested by the results of Kuiper [13] which show that if $n = 2$ imbeddings of class C^1 can have surprising properties.

Corollary 3 can be used to show that in the statements of some theorems, the requirement that a smooth surface be convex is superfluous. This point will be illustrated by a rigidity theorem of Pogorelov (cf. [18]).

Rigidity Theorem. Let S_1 be a 2-dimensional surface which bounds a convex body in E^3. Suppose S_1 has a spherical image 2π. Then, if S_2 is a convex surface isometric to S_1, S_2 is congruent to S_1.

If S_1 and S_2 are required to be of class C^2 and C^3 respectively and the isometry is of class C^2, it is not necessary to assume that S_2 is convex or even without self-intersections because these properties follow from Corollary 3.

Appendix 1. Sectional Curvature and the Second Fundamental Form.

The purpose of this section is to verify the proposition below which contains all of the assertions made above on the properties of the second fundamental form of a hypersurface which are determined intrinsically.

PROPOSITION. *Let M be a Riemannian n-manifold and $X: M \to E^{n+1}$ a C^2 isometric immersion of M. Let p be a point of M and let H be the matrix of coefficients of the second fundamental form of $X(M)$ at p. Let n_0, n_1, n_2 denote respectively the dimensions of the subspaces belonging to the zero, positive, and negative eigenvalues of H. If every sectional curvature is zero at p, then $n_0 \geqq n-1$. If at least one sectional curvature at p is not zero, then the sectional curvatures at p determine the numbers n_0, n_1, n_2 up to an interchange of n_1 and n_2.*

Note that the sectional curvatures are defined under the conditions of the proposition even though M may not be of class C^3 and the Riemannian-Christoffel tensor cannot be defined by the usual formulae; cf. [8], p. 912.

The proposition follows immediately from two lemmas which are stated below. Let $H = (h_{ij})$ be a real symmetric n by n matrix. Let V_n denote the space of all real n-vectors, and let $Rxyxy = (Hx, x)(Hy, y) - (Hx, y)^2$ for x, y in V_n. Let N denote the set of all n-vectors x such that $Rxyxy = 0$ for all y in V_n. Let N_0 denote the nullspace of H.

LEMMA 1A. *$N \supset N_0$. If $V_n = N$, then $\dim N_0 \geqq n-1$. If $V_n \neq N$, then $N = N_0$.*

Proof. The assertion $N \supset N_0$ follows immediately from the definition of $Rxyxy$. If $\dim N_0 < n-1$, there are two distinct unit orthogonal eigenvectors x, y of H which belong to the non-zero eigenvalues λ, μ of H. Then $Rxyxy = \lambda\mu \neq 0$, hence $N_0 \neq V_n$. Finally, suppose, if possible, that $V_n \neq N \neq N_0$. Then there must be a vector x in N orthogonal to N_0. In this case, H must be indefinite, for if H is semi-definite and $y \neq x$ is any vector orthogonal to N_0, the generalized Schwarz inequality gives $Rxyxy > 0$. Therefore H is indefinite and the subspaces N_1 and N_2 belonging respectively to the positive and negative eigenvalues of H are both non-empty.

Let $x = x_1 + x_2$, where x_i is in N_i. If x_1 and x_2 are both non-null, then $Rxyxy < 0$ for $y = x_1$. If x_1 is null, let y be any unit vector in N_1, and if x_2 is null, let y be any unit vector in N_2. In either case $Rxyxy < 0$. This shows that x is not in N which proves Lemma 1A.

If K is a subspace of the orthogonal complement of N having dimension

at least two, call K *definite* if for every pair of vectors x, y in K, $x \neq y$, $Rxyxy > 0$. Let $m = \operatorname{Max}\{\dim M : M \text{ is definite}\}$ (with $m = 0$ if there are no definite subspaces). Let $m_0 = \operatorname{Max}\{\dim N_1, \dim N_2\}$.

LEMMA 2A. *If $m = 0$, then $m_0 \leqq 1$. If $m \geqq 2$, $m = m_0$.*

Proof. If $m_0 \geqq 2$, $N = N_0$ by Lemma 1A, and N_1 and N_2 are definite subspaces. Therefore $m_0 \geqq 2$ implies $m \geqq m_0$. In particular if $m = 0$, then $m_0 \leqq 1$. If $m \geqq 2$, $N = N_0$ by Lemma 1A and again $m \geqq m_0$. Suppose that there is a definite subspace K such that $\dim K > \dim N_1$. Then K contains a vector $x \neq 0$ orthogonal to the projection of N_1 into K. Then x is in N_2. Similarly if $\dim K > \dim N_2$ there is a vector $y \neq 0$ in $K - N_2$. Then if $\dim K > m_0$ K contains non-null vectors x, y in N_2 and N_1 respectively. This implies $Rxyxy < 0$, hence K is not definite. This proves the last assertion of Lemma 2A.

Appendix 2. On the Extrema of the Cprvatures of a Surface.

1. The theorems of Hilbert and Weyl. Let S be a piece of two-dimensional surface of class C^2 in E^3. If p is a point on S, $k_1(p)$ and $k_2(p)$ will denote the principal curvatures of S at p, which are determined up to a factor ± 1. $H(p) = \frac{1}{2}(k_1 + k_2)$ and $K(p) = k_1 k_2$ will denote respectively the mean and Gaussian curvatures of S at p. S will be called *locally convex* if $K > 0$ everywhere on S and S has no self-intersections. In this case it will be supposed that the normal to S is directed in such a way that $H \geqq K^{\frac{1}{2}} > 0$ $k_1 \geqq k_2 > 0$. A function $f = f(p)$ defined on S will be said to have a *local maximum* [*minimum*] *at* p_0 if there is a neighborhood U of p_0 such that $f(p) \leqq f(p_0)$ [$f(p) \geqq f(p_0)$] for all p in U.

This appendix is concerned with the assertions:

(H_n) *Let S be a locally convex piece of surface of class C^n $(n \geqq 2)$. Suppose k_1 has a local maximum and k_2 a local minimum at a point p_0 on S. Then, in a neighborhood of p_0, S is a part of the surface of a sphere.*

(W_n) *Let S be a locally convex piece of surface of class C^n $(n \geqq 2)$. Suppose H has a local maximum and K a local minimum at a point p_0 on S. Then, in a neighborhood of p_0, S is a part of the surface of a sphere.*

The assertion (H_4) is due to Hilbert [9], Anhang V, p. 238, although he did not explicitly formulate (H_4). The proof fails if $n < 4$ because the existence and continuity of the second derivatives of k_1 and k_2 are used. Weyl proved (W_4) in [24], p. 72; cf. Chern [4], p. 287, for another proof.

Again both proofs fail if $n < 4$. Actually, (W_n) follows from (H_n). In fact, if H has maximum and K has a minimum at a point p_0 on S, then $k_1 = H + (H^2 - K)^{\frac{1}{2}}$ has a maximum and $k_2 = K/k_1$ a minimum at p_0. Therefore, a counterexample to (W_n) for any n is also a counterexample to (H_n).

Hilbert employed his theorem (H_4) to prove the rigidity of the sphere, and Chern used (H_4) to prove that all "special" Weingarten surfaces are spheres. A theorem of Grotemeyer [6] follows from (W_4) just as Chern's theorem follows from (H_4). Both Chern's and Grotemeyer's theorems are now known to be correct for surfaces of class C^2; cf. Pogorelov [19] and Aleksandrov [1]. In view of this, it is somewhat surprising that, as will be shown by the examples in Section 2 below,

(*) *the assertions* (H_2), (H_3), *and* (W_2) *are false.*

It will remain undecided whether or not the assertion (W_3) is correct.

2. Counterexamples. The counterexamples to (H_3) and (W_2) are both surfaces defined by functions of the form

$$z(x,y) = + ax^2 + y^2 - w(x,y,\lambda) + bw(y,x,\lambda) \tag{1}$$

for $x^2 + y^2 < R_0^2$, where

$$w(x,y,\lambda) = \tfrac{1}{2}x^2(x^2+y^2)^{\lambda}$$

is of class C^2 if $0 < \lambda \leqq \frac{1}{2}$ and of class C^3 if $\lambda > \frac{1}{2}$. R_0, a, b, λ are positive constants which will be specified more precisely later.

The curvatures of S are given by the formulae

$$\begin{aligned} H &= \tfrac{1}{2}\{(1+q^2)r - 2pqs + (1+p^2)t\}/(1+p^2+q^2)^{\frac{3}{2}}, \\ K &= (rt - s^2)/(1+p^2+q^2)^2, \end{aligned} \tag{2}$$

and

$$k_1, k_2 = H \pm (H^2 - K)^{\frac{1}{2}} \tag{3}$$

where, as usual, $p = z_x$, $q = z_y$, $r = z_{xx}$, $s = z_{xy}$, $t = z_{yy}$. It follows easily from (1)-(3) that the curvatures H, K, k_1, and k_2 are positive for $x^2 + y^2 < R_0^2$ if R_0 is sufficiently small. In order to determine whether the functions defined in (2), (3) have maxima or minina at the origin, their partial derivatives with respect to ρ for small $\rho > 0$ will be calculated. Here, (ρ, θ) are polar coordinates in the (x, y) plane.

A simple calculation shows that

$$p = 2a\rho\cos\theta + O(\rho^{2\lambda+1}), \qquad q = 2\rho\sin\theta + O(\rho^{2\lambda+1}) \tag{4}$$

and

$$r = 2a - \rho^{2\lambda} f_1(\theta,\lambda,b), \qquad s = O(\rho^{2\lambda}), \qquad t = 2 - \rho^{2\lambda} f_2(\theta,\lambda,b) \tag{5}$$

where the estimates hold as $\rho \to 0$ uniformly in θ, a, b, for $0 < a, b \leqq$ const., with λ fixed. The functions f_1 and f_2 are the trigonometric polynomials

$$\begin{aligned} f_1(\theta,\lambda,b) &= 1 + 5\lambda\cos^2\theta + 2\lambda(\lambda-1)\cos^4\theta \\ &\qquad - b\lambda\sin^2\theta(1+2(\lambda-1)\cos^2\theta) \\ f_2(\theta,\lambda,b) &= \lambda\cos^2\theta(1+2(\lambda-1)\sin^2\theta) \\ &\qquad - b(1+5\lambda\sin^2\theta + 2\lambda(\lambda-1)\sin^4\theta). \end{aligned} \tag{6}$$

Also, for $\rho \neq 0$,

$$p_\rho = 2a\cos\theta + O(\rho^{2\lambda}), \qquad q_\rho = 2\sin\theta + O(\rho^{2\lambda}) \tag{7}$$

and

$$r_\rho = -2\lambda\rho^{2\lambda-1} f_1(\theta,\lambda,b), \qquad s_\rho = O(\rho^{2\lambda-1}), \qquad t_\rho = -2\lambda\rho^{2\lambda-1} f_2(\theta,\lambda,b). \tag{8}$$

It is not difficult to see that, for $\rho \neq 0$, (2)-(8) imply

$$H_\rho = \tfrac{1}{2}(r_\rho + t_\rho) + O(\rho), \qquad K_\rho = 2(r_\rho + at_\rho) + O(\rho) + O(\rho^{4\lambda-1})$$

and for $a > 1$

$$k_{1\rho} = r_\rho + O(\rho) + O(\rho^{4\lambda-1}), \qquad k_{2\rho} = t_\rho + O(\rho) + O(\rho^{4\lambda-1}).$$

Hence

$$H_\rho = -\lambda\rho^{2\lambda-1}(f_1+f_2) + O(\rho) \tag{9}$$

$$K_\rho = -4\lambda\rho^{2\lambda-1}(f_1+af_2) + O(\rho) + O(\rho^{4\lambda-1}) \tag{10}$$

$$k_{1\rho} = -2\lambda\rho^{2\lambda-1}f_1 + O(\rho) + O(\rho^{4\lambda-1}) \tag{11}$$

$$k_{2\rho} = -2\lambda\rho^{2\lambda-1}f_2 + O(\rho) + O(\rho^{4\lambda-1}). \tag{12}$$

To obtain a counterexample to (W_2), let a, b be fixed, $0 < b < 1$, $ab > 1$. If $\lambda = 0$,

$$f_1 + f_2 \equiv 1 - b > 0 \text{ and } f_1 + af_2 = 1 - ab < 0.$$

The forms of f_1, f_2 show that if $\lambda = \lambda(a,b)$ is sufficiently small, then

$$f_1 + f_2 > 0 \text{ and } f_1 + af_2 < 0 \text{ for all } \theta. \tag{13}$$

It can be supposed that $\lambda < 1$. Then (9), (10), and (13) show that H has a relative maximum and K a relative minimum at the origin. A more detailed

computation shows that (13) cannot hold unless $\lambda < \frac{1}{2}$, hence a counterexample to (W_3) cannot be found in this manner.

A counterexample to (H_3) is obtained by choosing $a > 1$ and b and λ such that $\lambda > \frac{1}{2}$ and

$$f_1 > 0 \text{ and } f_2 < 0 \text{ for all } \theta. \tag{14}$$

Such a choice is possible because if $\lambda = \frac{1}{2}$ and $\frac{1}{2} < b < 2$ then $f_1 > 1 - b/2 > 0$ and $f_2 < \frac{1}{2} - b < 0$ for all θ. Then (11), (12), and (14) show that k_1 has a relative maximum and k_2 a relative minimum at the origin. This shows that (H_3) is false.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. D. Aleksandrov, "Uniqueness theorems for surfaces in the large," Part I, *Vestnik, Leningrad University*, 1956, no. 19, pp. 5-17; Part II, *ibid.*, 1957, no. 17, pp. 15-44; Part VI, *ibid.*, 1959, no. 1, pp. 5-13.

[2] N. Bourbaki, *Éléments de mathématique*, Livre V, *Espaces Vectoriels Topologiques*, Chapitres I, II, Paris, 1953.

[3] H. Buseman, *Convex Surfaces*, Interscience Publishers Inc. (1958).

[4] S. S. Chern and R. K. Lashof, "On the total curvature of immersed manifolds," *The Michigan Mathematical Journal*, vol. 5 (1958), pp. 5-12.

[5] S. S. Chern, "Some new characterizations of the Euclidean sphere," *Duke Mathematical Journal*, vol. 12 (1945), pp. 279-290.

[6] K. P. Grotemeyer, "Eine kennzeichende Eigenschaft der Kugel," *Archiv der Mathematik*, vol. 4 (1953), pp. 230-233.

[7] J. Hadamard, "Sur certaines propriétés des trajectoires en dynamique," *Journal de Mathématiques*, vol. 3 (1897), pp. 331-387.

[8] P. Hartman and L. Nirenberg, "On spherical image maps whose Jacobians do not change sign," *American Journal of Mathematics*, vol. 81 (1959), pp. 901-920.

[9] D. Hilbert, *Grundlagen der Geometrie*, 7th ed., Leipzig und Berlin, 1930.

[10] E. Hopf, "On S. Bernstein's theorem on surfaces $z(x, y)$ of nonpositive curvature," *Proceedings of the American Mathematics Society*, vol. 1 (1950), pp. 80-85.

[11] H. Hopf and W. Rinow, "Ueber die Begriff der vollständigen differentialgeometrischen Fläche," *Commentarii Mathematici Helvetici*, vol. 3 (1931), pp. 209-225.

[12] V. L. Klee, "Convex sets in linear spaces," *Duke Mathematical Journal*, vol. 18 (1951), pp. 443-465.

[13] N. H. Kuiper, "On C^1 isometric imbeddings," *Proceedings Koninglijke Nederlandse Akademie van Wetenschappen*, vol. 58 (1955), pp. 545-556 and pp. 683-689.

[14] A. Lichnerowicz, "Courbure, nombres de Betti, et espaces symétrique," *Proceedings of the International Congress of Mathematics*, vol. 2 (1950), pp. 216-223, Cambridge (Mass.).

[15] A. P. Morse, "The behavior of a function on its critical set," *Annals of Mathematics*, vol. 40 (1939), pp. 62-70.

[16] S. B. Myers, "Riemannian manifolds in the large," *Duke Mathematical Journal*, vol. 1 (1935) pp. 39-49.

[17] M. H. A. Newman, *The topology of plane sets*, 2nd ed., Cambridge University Press, 1954.

[18] A. V. Pogorelov, "On the rigidity of general convex surfaces with curvature 2π," *Doklady Akademie Nauk*, vol. 106 (1956), pp. 19-20 (in Russian).

[19] ———, "A general characteristic property of the sphere," *Uspekhi Matematicheskikh Nauk* (N. S.), vol. 11 (1956), no. 5(71), pp. 203-206 (in Russian).

[20] A. Pfluger, *Theorie der Riemannschen Flächen*, Springer-Verlag, Berlin, Göttingen, Heidelberg (1957).

[21] A. Sard, "The measure of the critical values of differentiable maps," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 883-890.

[22] J. J. Stoker, "Ueber die Gestalt der positiv gekrümmten offenen Flächen," *Compositio Mathematica*, vol. 3 (1936), pp. 55-88.

[23] J. Van Heijenoort, "On locally convex manifolds," *Communications on Pure and Applied Mathematics*, vol. 5 (1952), pp. 223-242.

[24] H. Weyl, "Ueber die Bestimmung einer geschlossen konvexen Fläche durch iher Linenelement," *Viertelijahrsschrift der Naturforschenden Gessellschaft in Zürich*, vol. 61 (1916), pp. 40-72.

[25] R. L. Wilder, *Topology of manifolds*, American Mathematical Society Colloquium Publications, vol. 32 (1949).

[26] G. T. Whyburn, *Analytic topology*, American Mathematical Society Colloquium Publications, vol. 28 (1942).

ON THE RATIONALITY OF THE ZETA FUNCTION OF AN ALGEBRAIC VARIETY.*

By Bernard Dwork.[1].

To Oscar Zariski on his sixtieth birthday.

Let p be a prime number, Ω the completion of the algebraic closure of the field of rational p-adic numbers and let $\mathfrak{K}$ be the residue class field of Ω. The field $\mathfrak{K}$ is the algebraic closure of its prime subfield and is of characteristic p. If T^* is the set of all roots of unity in Ω of order prime to p then the restriction of the residue class map to T^* is a multiplicative isomorphism of T^* onto the multiplicative group of $\mathfrak{K}$. The elements of $T = T^* \cup \{0\}$ form the Teichmüller representatives of $\mathfrak{K}$ in Ω and for each $x \in \mathfrak{K}$ the representative of x in Ω will be understood to be the element of T in the class x. The non-archimedean valuation of Ω will be denoted by the ordinal function, abbreviated "ord", and normalized by the condition $\operatorname{ord} p = 1$.

Let $\mathfrak{A}_n$ (resp: $\mathfrak{S}_n$) denote affine (resp: projective) space of dimension $n \geqq 1$ and characteristic p, viewed as consisting only of points with coordinates in $\mathfrak{K}$. Let k be the finite subfield of $\mathfrak{K}$ containing $q = p^a$ elements. A variety V in $\mathfrak{A}_n$ (resp: $\mathfrak{S}_n$) defined over k will be understood to be the subset consisting of the common zeros of a finite set of polynomials (resp: homogeneous polynomials) in n (resp: $n+1$) variables with coefficients in k. If V and W are varieties in $\mathfrak{A}_n$ (resp: $\mathfrak{S}_n$) defined over k then $V - W$ will be used to denote the set of points in V which are not in W and will be referred to as the difference between two varieties defined over k. Thus the empty subset of A_n (resp: S_n) is a variety and every variety is a difference between two varieties.

If V is the difference between two varieties defined over k, let N_s be the number of points of V with coordinates in the field of q^s elements in $\mathfrak{K}$. We define the zeta function of V to be the power series in one variable with rational coefficients:

* Received November 30, 1959.

[1] This work was supported in part by National Science Foundation Grant Number G7030.

$$\zeta(V;t) = \exp\{\sum_{s=1}^{\infty} N_s t^s/s\}. \tag{1}$$

It is an elementary consequence of Galois theory that the coefficients of this power series are integral. We note that while the expression "zeta function" has been used by other writers only in reference to non-singular, irreducible projective varieties, in our theory these restrictions play absolutely no role and therefore it serves no useful purpose in this discussion to restrict the expression in that way.

We have shown, [1], that the power series (1) represents an analytic function in the circle $\operatorname{ord} t > 0$ in Ω which has an analytic continuation to a meromorphic function in the region $\operatorname{ord} t > -\operatorname{ord} q$. This was shown to give some information about the zeta function in the classical sense (i. e., as a function on the complex plane), in particular that the first Betti number of a hypersurface of even dimension as defined by Weil [2] is certainly zero if the zeta function is rational.

In this paper we show that (1) always represents a rational function, thus generalizing and proving a part of Weil's conjecture [2]. This is done by showing (§ 3) that it represents a function having an analytic continuation to a meromorphic function on all of Ω. This gives rationality (§ 4) by arithmetic and function theoretic methods using the fact that the coefficients are rational integers and that the power series is (obviously) dominated in the archimedean sense by $(1-q^n t)^{-1}$.

The geometrical part of our treatment rests upon the computation of the number of points of a hypersurface rational over a finite field. The use of additive and multiplicative characters for such computations is classical and is closely connected with the theory of Gauss sums. The methods of Weil [2] and Hua and Vandiver [3] are readily generalized to arbitrary hypersurfaces giving explicit formulae in terms of Gauss sums. We avoid the use of Gauss sums as the Hasse-Davenport relations [4, equ. (0.8)] do not give enough information for general hypersurfaces. These relations, which play a key role in Weil's treatment of special hypersurfaces, are replaced here by a group theoretical device introduced previously, [1], and reproduced here (§ 2) so as to make this paper relatively self contained. In [1] we used only multiplicative characters, in particular an approximation of the trivial multiplicative character of finite fields suggested by Warning [5]. We now use approximations of the additive characters of finite fields. The main difficulty is the determination of multiplicative relations between additive characters of different finite fields (of the same characteristic). This difficulty is overcome (§ 1) by the construction of an Ω valued function θ on $\mathfrak{K}$ (or equivalently

on T) which can be used to split a non-trivial character of the field of p^s elements into a product of s factors. This splitting together with the group theoretical methods of § 2 gives the analytic continuation of the zeta function on Ω. Some connections betwen the θ function and the theory of Gauss sums are noted (§ 1).

I am indebted to Professor Washnitzer for many discussion of this problem and for considerable encouragement.

Throughout this paper, Ω, $\mathfrak{K}$, k, q, p, T, ord are as above. Z is the ring of integers, Z_+ the set of non-negative integers, Q the field of rational numbers, Q' the p-adic completion of Q in Ω, $\mathfrak{O}$ the ring of integers in Ω, $\mathfrak{O}'$ the ring of integers in Q' and $\mathfrak{o}$ the ring of p-integral rational numbers (i. e., $\mathfrak{o} = \mathfrak{O}' \cap Q$). If $X = (X_1, \cdots, X_n)$ is a set of indeterminates, $b \in Z$, $u = (u_1, \cdots, u_n) \in (Z_+)^n$ then X^b denotes the set $X_1{}^b, X_2{}^b, \cdots, X_n{}^b$ while X^u denotes the monomial $\prod_{i=1}^{n} X_i{}^{u_i}$. *Monomial* will always be used in this sense, i. e., the coefficient of a monomial will always be understood to be 1. If $u = (u_1, \cdots, u_n) \in Z_+{}^n$ then *weight* u is defined to be the real number $\sum_{i=1}^{n} u_i$.

1. The splitting of additive characters. Let L be the maximal unramified extension of Q' in Ω, σ the Frobenius automorphism of L over Q' and $Y_1, Y_2, \cdots, Y_m$ a finite set of variables. If $f(Y) = f(Y_1, \cdots, Y_m)$ is a power series in these variables with coefficients in L, let $f^\sigma(Y)$ be the power series obtained by replacing the coefficients of $f(Y)$ by their images under σ, let $f(Y^p) = f(Y_1{}^p, Y_2{}^p, \cdots, Y_m{}^p)$ and let $f(0) = f(0, 0, \cdots, 0)$.

LEMMA 1. *If $f(Y)$ is a power series in $Y_1, Y_2, \cdots, Y_m$ with coefficients in L and $f(0) = 1$ then the coefficients of f are integers in L if and only if the coefficients of the power series $f^\sigma(Y^p)/(f(Y))^p$ (other than the constant term) are divisible by p in the ring of integers of L.*

We omit the proof since this generalization of our criterion [6, Lemma 1] for p-adic integrality of coefficients of power series in one variable is proven precisely as in the case of one variable. The lemma will be applied to a power series in two variables with coefficients in Q and there will be no further reference to either the field L or the Frobenius automorphism.

If m is any non-negative integer, let $h(m, t)$ denote the binomial coefficient, $h(m, t) = (m!)^{-1} \prod_{j=0}^{m-1} (t - j)$, a polynomial in t of degree m with coefficients in Q. Let $H(t, Y) = (1 + Y)^t = \sum_{m=0}^{\infty} h(m, t) Y^m$, an element of

$Q\{t,y\}$, the ring of power series in two variables with rational coefficients. Let

$$F(t,Y)=H(t,Y)\prod_{j=1}^{\infty}H((t^{p^j}-t^{p^{j-1}})/p^j,Y^{p^j}), \tag{2}$$

an element of $Q\{t,Y\}$. From the definitions,

$$\begin{aligned}&F(t,Y)^p/F(t^p,Y^p)\\&\qquad=(1+Y)^{pt}/(1+Y^p)^t=\exp\{t\log[(1+Y)^p/(1+Y^p)]\}.\end{aligned}$$

It is well known that $\log(1+pt)\in pt\mathfrak{O}'\{t\}$ and that $\exp(pt)\in 1+pt\mathfrak{O}'\{t\}$ and thus we may conclude with the aid of the lemma since $(1+Y)^p/(1+Y^p)$ $\in 1+pY\mathfrak{o}\{Y\}$ that the coefficients of F are p-integral, i.e., $F\in\mathfrak{o}\{t,Y\}$.

Let $B_m(t)=\sum h(i_0,t)\prod_{j=1}^{m}h(i_j,(t^{p^j}-t^{p^{j-1}})/p^j)$, the sum being over all finite sequences $i_0,i_1,\cdots$ in Z_+ such that $i_0+pi_1+p^2i_2+\cdots=m$. Hence $B_m(t)$ is a sum of polynomials of degree m and therefore degree $B_m\leqq m$. It follows from the definitions that

$$F(t,Y)=\sum_{m=0}^{\infty}B_m(t)Y^m. \tag{3}$$

We now see that

$$F(t,Y)=\sum_{m=0}^{\infty}t^m\alpha_m(Y), \tag{4}$$

where $\alpha_m(Y)\in\mathfrak{o}\{Y\}$ and Y^m divides $\alpha_m(Y)$ in $\mathfrak{o}\{Y\}$. It is clear that as a power series in two variables $F(t,Y)$ converges in Ω for $\operatorname{ord}t\geqq 0$, $\operatorname{ord}Y>0$, that $\alpha_m(Y)$ converges under these conditions and that (3) and (4) converge to the same value. In particular, let $\lambda+1$ be a primitive p-th root of unity and let

$$\theta(t)=F(t,\lambda)=\sum_{m=0}^{\infty}\beta_m t^m \tag{5}$$

where β_m (sometimes denoted $\beta(m)$) $=\alpha_m(\lambda)$. We note that $\operatorname{ord}\lambda=1/(p-1)$ and therefore

$$\operatorname{ord}\beta_m\geqq m/(p-1), \tag{6}$$

an estimate of central importance in our computations.

For a fixed integer $s>1$, let t be a representative in T of an element t' in the field of p^s elements (hence $t^{p^s}=t$). It follows from (2), considering $F(t,Y)$ as a power series in Y with integral coefficients in $Q'(t)$, that

$$H(t+t^p+\cdots+t^{p^{s-1}},Y)=\prod_{i=0}^{s-1}F(t^{p^i},Y). \tag{7}$$

It is well known (and may be verified by means of the lemma) that if $x \in \mathfrak{O}'$ then $(1+Y)^x \in \mathfrak{O}'\{Y\}$ and therefore converges for $|Y| < 1$. It is easily verified that if L is the unramified extension of Q' of degree s (so the residue field of L is the field of p^s elements) and if S_s is the trace of L over Q' then $x \to H(S_s(x), \lambda)$ is a non-trivial additive character of the ring of integers of L which is trivial on the maximal ideal and therefore gives, by passage to quotients, a non-trivial additive character of the residue class field of L. Hence $t' \to \Theta_s(t') = H(S_s(t), \lambda) = H(t + t^p + \cdots + t^{p^{s-1}}, \lambda)$ is a non-trivial additive character of the field of p^s elements. Thus equation (7) may be written in the form

$$H(S_s(t), Y) = \prod_{i=0}^{s-1} F(t^{p^i}, Y) \tag{7'}$$

and replacing Y by λ we obtain

$$\Theta_s(t') = H(S_s(t), \lambda) = \prod_{i=0}^{s-1} \theta(t^{p^i}). \tag{8}$$

Thus (8) gives a splitting of Θ_s mentioned in the introduction.

Note. The object of the above discussion was the demonstration of the existence of a power series, $\theta(t)$, satisfying condition (6) and (8). These properties do not characterize $\theta(t)$ and certain remarks of the referee have led us to a somewhat simpler construction of a power series satisfying these conditions. Let $E(X)$ denote the Artin-Hasse exponential series,

$$E(X) = \exp\{-\sum_{j=0}^{\infty} X^{p^j}/p^j\}.$$

There exists a unique element, η, in $Q'(\lambda)$ such that $E(\eta) = 1 + \lambda$. It is an elementary exericse to verify that the power series $E(\eta t)$ satisfies conditions (6) and (8). This function is by no means new since it has appeared in investigations of the norm residue symbol of Kummer extensions of algebraic number fields.

Although not needed for our subsequent discussion, we note that (8) may be applied to the theory of Gauss sums. Let

$$j = j_0 + pj_1 + \cdots + p^{s-1}j_{s-1} \in Z, \qquad 0 \leqq j_i \leqq p-1,$$

not all j_i equal to $p-1$. Let $g_s(j) = \sum t^{-j}\Theta_s(t')$, the sum being over the $p^s - 1$ roots of unity in T. Then $g_s(j)$ is, as is well known [4], the image in Ω of a Gauss sum on the field of p^s elements. Using (5) and (8) we see that

$$g_s(j) = (p^s - 1) \sum \prod_{j=0}^{s-1} \beta(i_j), \tag{9}$$

the sum being over all $(i_0, i_1, \cdots, i_{s-1}) \in Z_+^s$ such that

$$\sum_{e=0}^{s-1} i_e p^e \equiv j \bmod p^s - 1.$$

From (2) we see that

$$\beta(m) = (-p^{-1}\log(1+\lambda^p))^m/m! \qquad \text{for } 0 \leqq m < p \tag{10}$$

and hence $\operatorname{ord}\beta(m) = m/(p-1)$ for $0 \leqq m < p$. It now follows from (6) that

$$g_s(j)/\prod_{i=0}^{s-1}\beta(j_i)) \equiv -1 \bmod \lambda^{p-1}. \tag{11}$$

Thus letting $\sigma(j) = j_0 + j_1 + \cdots + j_{s-1}$, $\gamma(j) = j_0! j_1! \cdots j_{s-1}!$, and noting that $p^{-1}\log(1+\lambda^p) \in p^{-1}\lambda^p(1+(\lambda^{p-1}))$, we have

$$-g_s(j)/(-\lambda^p/p)^{\sigma(j)} \equiv \gamma(j)^{-1} \bmod \lambda^{p-1}. \tag{12}$$

Stickelberger's congruence [17],

$$-g_s(j)/\lambda^{\sigma(j)} \equiv \gamma(j)^{-1} \bmod \lambda \tag{12.1}$$

follows directly from (12) since

$$\lambda^{p-1}/(-p) \equiv 1 \bmod \lambda. \tag{12.2}$$

While (12) is ostensibly stronger than (12.1), it is in fact a consequence of (12.1) and the fact that $Q'(\lambda)$ is a Kummer extension of Q'. It follows from (12.2) that there exists, π, a root of $x^{p-1} + p = 0$ in $Q'(\lambda)$ such that $\lambda/\pi \equiv 1 \bmod \lambda$. Hence, letting $u = -g_s(j)/(\pi^{\sigma(j)}/\gamma(j))$, it follows from (12.1) that $u \equiv 1 \bmod \lambda$. If α is any automorphism of $Q'(\lambda)/Q'$ then $u^{1-\alpha}$ is a $p-1$ root of unity and also $u^{1-\alpha} \equiv 1 \bmod \lambda$. Hence $u^{1-\alpha} = 1$ which shows that $u \in Q' \cap (1+(\lambda)) = 1 + (p) \subset 1 + (\lambda^{p-1})$. Finally

$$\lambda^p/(-p\pi) = \lambda^p/\pi^p \in (1+(\lambda))^p \subset 1 + (\lambda^p).$$

Hence $u(-p\pi/\lambda^p)^{\sigma(j)} \equiv 1 \bmod \lambda^{p-1}$ which is equivalent to (12).

2. Linear transformations of polynomial rings. In this section we describe a group theoretical device discussed in greater detail in [1].

Let $L[X] = L[X_1, \cdots, X_n]$ be the ring of polynomials in n variables over an arbitrary field, L. Let ψ be the endomorphism of $L[X]$ (as L-module, not as ring) defined by

$$\psi(X^u) = \begin{cases} X^{u/q} & \text{if } q \mid u \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

for all $u \in Z_+^n$. (In this section q need not be a power of a prime. In the application (§ 3), $q = p^a$). For $H \in L[X]$, let $\psi \circ H$ denote the endomorphism $\xi \to \psi(H\xi)$ of $L[X]$. For each $m \in Z_+$, let L_m denote the finite dimensional subspace of $L[X]$ consisting of all polynomials of degree not greater than m and let $(\psi \circ H)_m$ be the restriction of $\psi \circ H$ to L_m. It is easily verified that for $m \geqq m_0 = (\text{degree}\, H)/(q-1)$,

(i) $(\psi \circ H)_m$ is an endomorphism of L_m

(ii) the "characteristic polynomial," $\det(I - t(\psi \circ H)_m)$ is independent of m

(iii) for each integer s, $s \geqq 1$, the trace, $\mathrm{Tr}((\psi \circ H)_m)^s$ is independent of m.

We are therefore able to define $\det(I - t(\psi \circ H))$ and $\mathrm{Tr}(\psi \circ H)^s$ in a natural way.

Let now L be algebraically closed and of characteristic zero. Let G_1 be the group of all roots of unity in L of order prime to q and let G be (for some fixed integer $n \geqq 1$) the n-fold direct product of G_1. Our technical device may now be stated:

LEMMA 2. *If* $H \in L[X] = L[X_1, \cdots, X_n]$, $s \in Z$, $s \geqq 1$, *then*

$$(q^s - 1)^n \,\mathrm{Tr}(\psi \circ H)^s = \sum H(x) H(x^q) \cdots H(x^{q^{s-1}}), \tag{14}$$

the sum being over all $x \in G$ *such that* $x^{q^s - 1} = 1$.

Proof. Since $(\psi \circ H)^s = \psi^s \circ \{H(X) H(X^q) \cdots H(X^{q^{s-1}})\}$ and ψ^s is the endomorphism of $L[X]$ obtained by replacing q by q^s in the definition of ψ, we see that by taking a new value for q and a new choice for H we may assume $s = 1$. But $H \to \mathrm{Tr}(\psi \circ H)$ and $H \to \sum H(X)$ (the sum being over the elements of G of exponent $q-1$) are homomorphisms of $L[X]$, as L module, into L. Hence it may be assumed that H is a monomial and the verification becomes trivial: if $H = X^v$ then we find $\mathrm{Tr}(\psi \circ H) = 1$ if $(q-1) \mid v$ and is zero otherwise.

3. The meromorphic character of the zeta function. Let V be the difference between two varieties defined over k (in either $\mathfrak{A}_n$ or $\mathfrak{S}_n$). We now show that the analytic function in the circle $\mathrm{ord}\, t > 0$ in Ω represented by (1) has an analytic continuation to a meromorphic function on all of Ω, i.e., we show that the power series is the ratio of two power series in $1 + t\Omega\{t\}$ which converge for all $t \in \Omega$. While Krasner [8] has developed a theory of

analytic continuation in Ω, this will not be needed since the equivalent statements in terms of formal power series will be adequate for our purpose.

THEOREM 1. *$\zeta(V,t)$ is meromorphic.*

Proof. 1. To fix ideas let V be a difference between two varieties, say $V_1 - V_2$ which lie in $\mathfrak{A}_n$ (the projective case requires no more than changes in notation). Then $V = V_1 - V_1 \cap V_2$ and therefore

$$\zeta(V,t) = \zeta(V_1,t)/\zeta(V_1 \cap V_2, t)$$

Hence it is enough to prove the assertion for varieties defined over k, i.e., let $V = \bigcap_{i=1}^{r} V_i$, V_i a hypersurface defined over k. Let S_r be the set $\{1,2,\cdots,r\}$ and for each non-empty subset, B, of S_r let $W_B = \bigcup_{i\in B} V_i$. Then W_B is a hypersurface and by an obvious combinatorial argument,

$$\zeta(V,t) = \prod \zeta(W_B,t)^{-(-1)^{m(B)}}, \tag{15}$$

where $m(B)$ is the number of elements in B and the product is over all subsets, B, of S_r. Hence it is enough to prove the assertion for hypersurfaces.

Let V be a hypersurface in $\mathfrak{A}_n$, let S be the set $\{1,2,\cdots,n\}$. For each (proper or improper) subset B of S, let B' be the complementary subset of S, let W_B be the linear subvariety $\{X_i = 0\}_{i\in B}$ and let U_B be the degenerate hypersurface: $\prod_{i\in B} X_i = 0$. (If B is the empty subset of S, we understand U_B to be the empty subset of $\mathfrak{A}_n$ and W_B to be $\mathfrak{A}_n$). Clearly,

$$V = \cup (V \cap W_B - V \cap U_{B'}),$$

a disjoint union indexed by the subsets B of S. Hence it is enough to show that $Z(V - U_S, t)$ is meromorphic. This completes our reduction process.

2. Let $f(X) \in k[X] = k[X_1,\cdots,X_n]$ be the defining polynomial of a hypersurface V in $\mathfrak{A}_n$. Let V' be the degenerate hypersurface: $\prod_{i=1}^{n} X_i = 0$. We compute N_s, the number of points of $V - V'$ which are rational over k_s, the field of q^s elements in K.

Let Θ be a non-trivial additive character of k_s. Since $q = p^a$, we may take Θ to be Θ_{as} in the notation of §1. For $u \in k_s$, $\sum \Theta(ux_0) = q^s$ if $u = 0$, zero otherwise, the sum being over all $x_0 \in k_s$. We write the sum in the form $1 + \sum \Theta(ux_0)$ the sum now being over the multiplicative group, k_s^* of k_s. Hence

$$q^s N_s = (q^s - 1)^n + \sum \Theta(x_0 f(x)) \tag{16}$$

the sum being over all $x \in k_s^{*n}$, $x_0 \in k_s^*$. It is now convenient to represent the polynomial $X_0 f(X) = X_0 f(X_1, \cdots, X_n)$ explicitly as a sum, $\sum_{i=1}^{\rho} A_i M_i$, where $A_1, A_2, \cdots, A_\rho$ are elements of k^* and $M_1, M_2, \cdots, M_\rho$ is a set of monomials in $n+1$ variables. Specifically, $M_i = X^{w_i}$ where $\{w_i\}_{i=1}{}^{\rho}$ is a set of ρ distinct elements of Z_+^{n+1}, it being understood that X now denotes the variables $X_0, X_1, \cdots, X_n$. We note (without further comment) that the first coefficient of each of the w_i is 1. Using the additive property of $\circledast$, (16) becomes

$$q^s N_s = (q^s - 1)^n + \sum \prod_{i=1}^{\rho} \circledast(A_i M_i), \tag{17}$$

the sum being over all $x \in k_s^{*n+1}$.

For $i = 1, 2, \cdots, \rho$, A_i has a Teichmüller representative in Ω again denoted by A_i. Let T_s denote the $q^s - 1$ roots of unity in Ω, i.e., the Teichmüller representatives of k_s^*. Since $q = p^a$, it follows from § 1 that

$$q^s N_s = (q^s - 1)^n + \sum \prod_{i=1}^{\rho} \prod_{j=0}^{as-1} \theta((A_i M_i)^{p^j}) \tag{18}$$

the sum now being over the $(n+1)$-fold direct product T_s^{n+1} of T_s. Let $\Lambda(t) = \prod_{i=0}^{a-1} \theta(t^{p^i})$. Then $\Lambda(t) = \sum_{m=0}^{\infty} \lambda_m t^m$, where $\lambda_0 = 1$ and in general (writing $\beta(i)$ for β_i as in § 1), $\lambda_m = \sum \beta(i_0)\beta(i_1) \cdots \beta(i_{a-1})$, the sum being over all $(i_0, i_1, \cdots, i_{a-1}) \in Z_+^a$ such that $m = \sum_{j=0}^{a-1} i_j p^j$. To estimate $\operatorname{ord} \lambda_m$, we note that

$$\begin{aligned} \operatorname{ord}\Big(\prod_{j=0}^{a-1} \beta(i_j)\Big) &\geqq (i_0 + i_1 + \cdots + i_{a-1})/(p-1) \\ &\geqq mp/(q(p-1)) \geqq m/(q-1). \end{aligned}$$

Hence

$$\operatorname{ord} \lambda_m \geqq m/(q-1) \tag{6'}$$

and furthermore $\prod_{j=0}^{as-1} \theta(t^{p^j}) = \prod_{j=0}^{s-1} \Lambda(t^{q^j})$. Since $A_i{}^q = A_i$, $i = 1, 2, \cdots, \rho$, equation (18) assumes the form

$$q^s N_s = (q^s - 1)^m + \sum \prod_{i=1}^{\rho} \prod_{j=0}^{s-1} \Lambda(A_i M_i{}^{q^j}), \tag{18'}$$

the sum being as in (18).

It would be desirable to apply the methods of § 2 to the right side of (18′). Since no formulation of § 2 in terms of infinite series is available we proceed by means of p-adic approximations. For $r \in Z$, $r \geqq 1$, let

$$\Lambda_r(t) = \sum_{i=0}^{r(q-1)} \lambda_i t^i$$

so that for ord $t \geqq 0$,

$$\Lambda_r(t) \equiv \Lambda(t) \bmod p^r.$$

Let

$$F_r(X) = \prod_{i=1}^{\rho} \Lambda_r(A_i M_i),$$

then

$$\prod_{j=0}^{s-1} F_r(X^{q^j}) = \prod_{i=1}^{\rho} \prod_{j=0}^{s-1} \Lambda_r(A_i M_i^{q^j}) \equiv \prod_{i=1}^{\rho} \prod_{j=0}^{s-1} \Lambda(A_i M_i^{q^j}) \bmod p^r.$$

Hence

$$q^s N_s \equiv (q^s - 1)^n + \sum \prod_{j=0}^{s-1} F_r(X^{q^j}) \bmod p^r, \tag{19}$$

the sum being as in (18). Thus from Lemma 2,

$$q^s N_s \equiv (q^s - 1)^n + (q^s - 1)^{n+1} \operatorname{Tr}(\psi \circ F_r)^s \bmod p^r. \tag{20}$$

Hence for each integer r greater than 1, there exists a sequence of elements of Ω, $\{b_{r,s}\}_{s=0}^{\infty}$ such that ord $b_{r,s} \geqq r$ and such that

$$q^s N_s = (q^s - 1)^n + (q^s - 1)^{n+1} \operatorname{Tr}(\psi \circ F_r)^s + b_{r,s} \tag{20.1}$$

In the group $1 + t\Omega\{t\}$ of formal power series in one variable with coefficients in Ω and constant term 1, let the *weak topology* be the topology of pointwise convergence of coefficients, i. e., the multiplicative groups $V(m, \alpha) = 1 + t^m\Omega\{t\} + \alpha t\mathfrak{O}\{t\}$, $\alpha \in \mathfrak{O}$, $m \in Z_+$, $m > 0$, form a basis of the neighborhoods of 1. We note that $1 + t\mathfrak{O}\{t\}$ is a complete topological group under the weak topology. Let δ be the homomorphism $h(t) \to h(t)/h(tq)$ of $1 + t\Omega\{t\}$ into itself. Clearly $\delta^{-1}h(t) = \prod_{i=1}^{\infty} h(tq^i)$, the product being convergent in the weak topology, and so δ is a group automorphism of $1 + t\Omega\{t\}$. On the other hand δ and δ^{-1} map $V(m, \alpha)$ into itself for each $\alpha \in \mathfrak{O}$ and each integer, m, greater than zero. Hence δ is a topological group automorphism of $1 + t\Omega\{t\}$.

Let ϕ be the mapping $g(t) \to g(tq)$ of $\Omega\{t\}$ into itself. Clearly

$$\log g(t)^{\delta} = (1 - \phi) \log g(t)$$

where 1 denotes the identity mapping of $\Omega\{t\}$ into itself. Hence

$$\sum_{s=1}^{\infty} (q^s - 1)^n t^s / s = -(\phi - 1)^n \log(1 - t) = \log(1 - t)^{-(-\delta)^n}$$

and

$$\sum_{s=1}^{\infty}(q^s-1)^{n+1}(t^s/s)\operatorname{Tr}(\psi\circ F_r)^s = -(\delta-1)^{n+1}\log\det(I-t\psi\circ F_r)$$
$$=\log\det(I-t\psi\circ F_r)^{-(-\delta)^{n+1}}.$$

It follows from (20.1) and these purely formal (i.e. non-topological) properties of the δ operator that

$$(20.2)\qquad \zeta(V-V',qt)=(1-t)^{-(-\delta)^n}\det\,(I-t\psi\circ F_r)^{-(-\delta)^{n+1}}\exp\{\sum_{s=1}^{\infty}b_{r,s}t^s/s\}.$$

It is easily verified that for $\operatorname{ord}\alpha\geqq 1/(p-1)$, $\exp\{V(m,\alpha)-1\}\subset V(m,\alpha)$, i.e., if we define the weak topology on $t\Omega\{t\}$ in the obvious way, then $g(t)\to\exp\{g(t)\}$ is a continuous map of $t\Omega\{t\}$ into $1+t\Omega\{t\}$. In the weak topology of $t\Omega\{t\}$, we have $\lim\limits_{r\to\infty}\sum\limits_{s=1}^{\infty}b_{r,s}t^s/s=0$ and hence

$$\lim_{r\to\infty}\exp\{\sum_{s=1}^{\infty}b_{r,s}t^s/s\}=1.$$

We may now conclude from (20.2) and the topological group theoretic properties of the δ operator that $\Delta(t)=\lim\limits_{r\to\infty}\det(I-t\psi\circ F_r)$ exists in our topology and that

$$(21)\qquad \zeta(V-V',qt)=(1-t)^{-(-\delta)^n}\Delta(t)^{-(-\delta)^{n+1}}.$$

Let $\det(I-t\psi\circ F_r)=\sum\limits_{m=0}^{\infty}\gamma_{r,m}t^m$. We shall show that there exist non-negative real numbers, $z_1,z_2,\cdots$ such that $z_m\to\infty$ and $(\operatorname{ord}\gamma_{r,m})/m\geqq z_m$ for all $r,m\in Z_+$. It will then be clear that $\Delta(t)=\lim\limits_{r\to\infty}\det(I-t\psi\circ F_r)$ is a power series in $\Omega\{t\}$ which converges for all $t\in\Omega$, i.e., is entire. This together with (21) will complete the proof of the theorem.

Let d be the degree of the (not necessarily homogeneous) polynomial f in n variables, defining the hypersurface V. We write $F_r(X)=\sum B_uX^u$, the sum being over some finite subset of Z_+^{n+1}. It follows from the definitions that

$$(22)\qquad B_u=\sum\prod_{i=1}^{\rho}\lambda(b_i)A_i^{b_i}$$

the sum being over all $(b_1,b_2,\cdots,b_\rho)\in Z_+^{\rho}$ such that

$$(23)\qquad \sum_{i=1}^{\rho}b_iw_i=u,$$

it being understood in (22) that $\lambda(m)$ denotes λ_m. Since weight $w_i\leqq d+1$ for $i=1,2,\cdots,\rho$, it follows from (23) that

$$\text{weight } u \leqq (d+1)\sum_{i=1}^{\rho} b_i \leqq (d+1)(q-1)\operatorname{ord}\left(\prod_{i=1}^{\rho}\lambda(b_i)A_i^{b_i}\right)$$

Hence

$$\operatorname{ord} B_u \geqq (\text{weight } u)/((d+1)(q-1)), \tag{24}$$

an estimate independent of r. For convenience let B_u also be denoted by $B(u)$. For fixed r, we may form a matrix representation, E, of $\psi \circ F_r$, indexed by a finite set of pairs (u,v) of elements of Z_+^{n+1}, by letting $E(u,v)$ be the coefficient of X^v in the polynomial $\psi(X^u F_r(X))$. Clearly $E(u,v) = B(qv-u)$. But $\gamma_{r,m}$ is the coefficient of t^m in the "characteristic equation" $\det(I-tE)$ of E and therefore is a sum and difference of products of the form

$$P = \prod_{i=1}^{m} E_r(u_i, v_i),$$

where $\{u_1, u_2, \cdots, u_m\}$ is a set of m *distinct* elements of Z_+^{n+1} and $\{v_1, v_2, \cdots, v_m\}$ is the same set in a possibly different order. Using (24) we obtain, $(d+1)(q-1)\operatorname{ord} P \geqq \sum_{i=1}^{m}\text{weight}(qv_i-u_i) = (q-1)\sum_{i=1}^{m}\text{weight } u_i$. There are only $\binom{n+d}{d}$ elements of Z_+^{n+1} of weight d. For each $m \in Z_+$ there exists a unique $x \in Z_+$ such that

$$m = D_m + \sum_{i=0}^{x}\binom{n+i}{i} \tag{25}$$

where $0 \leqq D_m < \binom{n+x+1}{x+1}$. We now have by the distinctness of the u_i, $(d+1)\operatorname{ord} P \geqq \sum_{i=0}^{x} i\binom{n+i}{i} + (x+1)D_m$ and therefore

$$(d+1)\operatorname{ord}\gamma_{r,m} \geqq \sum_{i=0}^{x} i\binom{n+i}{i} + (x+1)D_m, \tag{26}$$

an estimate which is independent of r. Let $z_m(d+1)m$ be the right side of (26), then z_m is independent of r, and $(\operatorname{ord}\gamma_{r,m})/m \geqq z_m$ for all $r, m \in Z_+$. We claim that $\lim_{m\to\infty} z_m = \infty$. Let $a_i = \binom{n+i}{i}$ and note that $a_i \leqq a_{i+1}$. Clearly $(d+1)z_m \leqq x+1$ and hence

$$(d+1)z_m \geqq \left(\sum_{i=0}^{x} ia_i\right)/\sum_{i=0}^{x} a_i \geqq \sum_{i=[x/2]}^{x} ia_i/\sum_{i=0}^{x} a_i \geqq \tfrac{1}{2}\sum_{i=[x/2]}^{x} ia_i/\sum_{i=[x/2]}^{x} a_i \geqq [x/2]/2,$$

and furthermore equation (25) shows that $x\to\infty$ as $m\to\infty$. Hence $z_m\to\infty$ as asserted and this completes the proof of the theorem.

4. Borel's theorem. In this section we complete the proof of rationality by function theoretic methods. In addition to our previous conventions we introduce the following: $|\ |$ will be used to denote the ordinary absolute value of real and complex numbers, $|\ |_p$ will be used for the valuation in Ω, normalized so that $|p|_p = 1/p$. The determinant of an $m \times m$ matrix will be indicated by given a typical line $\| a_{i,1}, a_{i,2}, \cdots, a_{i,m} \|_{i=1}{}^m$.

We shall make considerable use of the methods of Borel [9]. In particular we shall use

(1) A power series with integral coefficients which is meromorphic in a circle of radius greater than one in the complex plane represents a rational function.

(2) If $F(t) = \sum_{s=0}^{\infty} A_s t^s$ is a formal power series with coefficients in any field, let

$$N_{s,m} = \| A_{s+j}, A_{s+j+1}, \cdots, A_{s+j+m} \|_{j=0}{}^m, \tag{27}$$

then $F(t)$ is certainly a ratio of polynomials if there exists an integer, m, such that $N_{s,m} = 0$ for all integers, s, which are large enough.

We pause to state some known facts about functions on Ω, [10]. Everything stated may be deduced easily from the theory of Newton polygons of power series. To the best of our knowledge no proof of this theory is available in the literature. To overcome this deficiency an exposition of this theory together with proofs of the next two propositions will be given in a future paper. For the present we shall state what is needed and indicate an alternate treatment adequate to complete the proof of rationality of the zeta function.

For each element b of the extended real line, $[-\infty, \infty)$ let

$$U_b = \{x \in \Omega \mid \operatorname{ord} x > b\}.$$

An element

$$F(t) = \sum_{s=0}^{\infty} A_s t^s$$

of $\Omega\{t\}$ which converges in U_b is said to represent an *analytic function* in that region. If F is the ratio F_1/F_2 of power series F_1, F_2 which converge in U_b then F is said to represent a *meromorphic function* on U_b. F is said to represent an *entire* function if it converges on $U_{-\infty} = \Omega$.

PROPOSITION 1. *If F converges in $U_{b'}$ and is never zero on U_b, where $-\infty \leqq b' < b < \infty$ then the power series $1/F$ converges in U_b.*

PROPOSITION 2. *If F converges in $U_{b'}$ and $-\infty \leqq b' < b < \infty$ then there*

exists a polynomial, $P(t)$, and an element, $H(t) \in \Omega\{t\}$, (both depending upon b) such that H converges in $U_{b'}$, is never zero in U_b and such that $F(t) = P(t)H(t)$.

The following direct consequence of these statements is needed for our generalization of Borel's Theorem (Theorems 2 and 3 below).

PROPOSITION 3. *If F represents a function meromorphic on $U_{b'}$ and $-\infty \leqq b' < b < \infty$ then there exists a polynomial, $P(t)$, depending upon b, such that the power series $P(t)F(t)$ converges in U_b.*

For the proof of Theorem 2 (and hence to complete the proof of rationality of the zeta function) it is enough to know Proposition 3 for the special case $-\infty = b'$. In this case the statement of Proposition 3 is a direct consequence of a theorem of Schnirelmann [11]:

A power series in one variable which converges everywhere in Ω (i. e. an entire function) must be of the form

$$\alpha t^m \prod_{i=1}^{\infty} (1 - \lambda_i t)$$

where $\alpha, \lambda_1, \lambda_2, \cdots \in \Omega$, $m \in Z_+$ and $\lim_{i \to \infty} \lambda_i = 0$.

We note that this theorem is itself an elementary consequence of the previously mentioned theory of Newton polygons.

THEOREM 2. *If $F(t) = \sum_{s=0}^{\infty} A_s t^s \in Z\{t\}$ converges in the complex plane in a circle of radius R and is meromorphic in Ω in the circle, $|t|_p < r$, where*

$$Rr > 1$$

then F represents a rational function.

Proof. In the following, the symbols R and r are used to denote the radii of circles properly contained by the circles in the statement of the theorem, but so chosen that the inequality, $Rr > 1$, remains valid.

In view of Borel's theorem, we may assume that $R \leqq 1$ and therefore $r > 1$. By hypothesis there exists a polynomial, $P(t) = 1 + a_1 t + \cdots + a_e t^e$, such that $P(t)F(t) = \sum_{s=0}^{\infty} B_s t^s$ converges in Ω for $|t|_p \leqq r$. Hence there exists a positive real number, M, such that

$$(28) \qquad |A_s| < M/R^s, \qquad |B_s|_p < M/r^s.$$

Clearly, $B_{s+e} = A_{s+e} + a_1 A_{s+e-1} + \cdots + a_e A_s$, and therefore for $m > e$ we have

(29) $$N_{s,m} = \| A_{s+j}, A_{s+j+1}, \cdots, A_{s+j+e-1}, B_{s+j+e}, \cdots, B_{s+j+m} \|_{j=0}{}^{m},$$

and since $|A_s|_p \leqq 1$, $r > 1$,

(30) $$|N_{s,m}|_p \leqq M^{m+1-e}/r^{s(m+1-e)}.$$

Using (27) and (28), we have

(31) $$|N_{s,m}| \leqq (m+1)!\,M^{m+1}/R^{(s+2m)(m+1)}.$$

By hypothesis we may choose a positive integer, m, so large that $R^{m+1}r^{m+1-e} > 1$. For this value of m, it follows from (30) and (31) that there exists $s_0 \in Z_+$ such that $|N_{s,m}|_p \, |N_{s,m}| < 1$ for $s > s_0$. Since $N_{s,m} \in Z$, we see that $N_{s,m}$ does not satisfy the product formula for valuations in Q and therefore must be zero for $s > s_0$. The rationality of F now follows from the second result of Borel.

This completes the proof of rationality of $\zeta(V,t)$, since in the notation of § 3 the power series is dominated in the ordinary sense by $(1 - q^n t)^{-1}$ and therefore has a non-zero radius of convergence in the complex plane.

A further generalization of Borel's theorem is worth noting. If L is an algebraic number field and $\mathfrak{p}$ is a prime of L (finite or infinite) let $\Omega_{\mathfrak{p}}$ be the completion of the algebraic closure of the completion of L at $\mathfrak{p}$. We normalize the valuation $|\ |_{\mathfrak{p}}$ of $\Omega_{\mathfrak{p}}$ so that the product formula, $\prod |\alpha|_{\mathfrak{p}} = 1$, holds for all non-zero elements, α, of L, the product being over all primes of L. In particular if $\mathfrak{p}$ is an infinite prime then $\Omega_{\mathfrak{p}}$ is the field of complex numbers and $|\ |_{\mathfrak{p}}$ is the ordinary (resp: square of the ordinary) absolute value in that field if $\mathfrak{p}$ is a real (resp: complex) prime of L.

THEOREM 3. *If L is an algebraic number field and $F(t) = \sum_{i=1}^{\infty} A_s t^s \in L\{t\}$, then F is rational if and only if there exists a finite set, S, of primes of L such that*

(i) For each $\mathfrak{p} \notin S$, $|A_s|_{\mathfrak{p}} \leqq 1$ for all non-negative integers s.

(ii) For each $\mathfrak{p} \in S$, $F(t)$ is meromorphic in $\Omega_{\mathfrak{p}}$ in a circle $|t|_{\mathfrak{p}} \leqq R_{\mathfrak{p}}$, where $\{R_{\mathfrak{p}}\}_{\mathfrak{p} \in S}$ is a set of positive real numbers satisfying the condition

(32) $$\prod_{\mathfrak{p} \in S} R_{\mathfrak{p}} > 1.$$

Proof. If F is rational then (ii) is certainly satisfied if S is any non-empty set of primes. On the other hand if F is a ratio of polynomials with coefficients in a field containing L then since $F \in L\{t\}$, it is a ratio $f(t)/g(t)$

of polynomials, f, g, with coefficients in L and hence with no loss in generality, it may be assumed that the coefficients of f and g are algebraic integers and that $g(0) \neq 0$. Hence the coefficients of F are integral at each finite prime of L which does not divide $g(0)$. Hence (i) is satisfied if we take S to be the set of all infinite primes of L and all prime divisors of $g(0)$. (In particular if L is the field of rational numbers then (i) is a consequence of Eisenstein's Theorem: If $F \in Q\{t\}$ and represents algebraic function then $F \in Z\{t/m\}$ for some integer m.)

To prove the "if" part of the theorem we repeat the argument of Theorem 2. With no loss in generality we may supppose that the infinite primes of L lie in S and that for each $\mathfrak{p} \in S$, F is meromorphic in a circle of radius *strictly greater* than $R_{\mathfrak{p}}$ and that inequality (32) is still valid. Since the radius of convergence of F is non-zero at each prime of L, there exists a real number $D > 0$, such that for each prime $\mathfrak{p} \in S$, the radius of convergence of F at $\mathfrak{p}$ is strictly greater than D. If $\mathfrak{p} \in S$ then there exists a polynomial $P_{\mathfrak{p}}(t) = 1 + a_1 t + \cdots + a_e t^e$ such that $P_{\mathfrak{p}}(t)F(t) = \sum_{s=0}^{\infty} B_{s,\mathfrak{p}} t^s$ converges in $\Omega_{\mathfrak{p}}$ in a circle of radius strictly greater than $R_{\mathfrak{p}}$. The coefficients, $a_1, a_2, \cdots, a_e$ of $P_{\mathfrak{p}}(t)$ lie in $\Omega_{\mathfrak{p}}$ and depend upon $\mathfrak{p}$ but it may be assumed that e is independent of $\mathfrak{p}$, that is, $e \geqq \deg P_{\mathfrak{p}}$ for all $\mathfrak{p} \in S$. Since S is a finite set, there exists a real number M such that

$$(33) \qquad |B_{s,\mathfrak{p}}|_{\mathfrak{p}} \leqq M/R_{\mathfrak{p}}^{s}, \qquad |A_s|_{\mathfrak{p}} \leqq M/D^s$$

for each $\mathfrak{p} \in S$ and each non-negative integer, s. For each $\mathfrak{p} \in S$ equation (29) is valid if B_s is replaced by $B_{s,\mathfrak{p}}$. Let $z = (m + e - 1)e$ if $D < 1$, $z = 0$ if $D \geqq 1$. For $\mathfrak{p} \in S$, let $\mu(\mathfrak{p}) = 2m(m + 1 - e)$ if $R_{\mathfrak{p}} < 1$, $\mu(\mathfrak{p}) = 0$ if $R_{\mathfrak{p}} \geqq 1$. It follows from (33) and (29) that for $\mathfrak{p} \in S$,

$$(34) \qquad |N_{s,m}|_{\mathfrak{p}} \leqq (m+1)!\, M^{m+1}/\{D^{se+z} R_{\mathfrak{p}}^{s(m+1-e)+\mu(\mathfrak{p})}\}$$

and therefore

$$(35) \qquad \prod_{\mathfrak{p} \in S} |N_{s,m}|_{\mathfrak{p}} \leqq G_m/(\prod_{\mathfrak{p} \in S} D^e R_{\mathfrak{p}}^{m+1-e})^s$$

where $G_m = \prod_{\mathfrak{p} \in S} \{(m+1)!\, M^{m+1}/D^z R_{\mathfrak{p}}^{\mu(\mathfrak{p})}\}$ is a real number depending upon m but independent of s. Using (32) we see that m may be chosen such that

$$\prod_{\mathfrak{p} \in S} \{D^e R_{\mathfrak{p}}^{m+1-e}\} > 1.$$

Let m be so chosen, then $\prod_{\mathfrak{p} \in S} |N_{s,m}|_{\mathfrak{p}} < 1$ for all s greater than some integer s_0. Since $|N_{s,m}|_{\mathfrak{p}} \leqq 1$ for $\mathfrak{p} \notin S$, it is clear that $N_{s,m}$ does not satisfy the product

formula in L for $s > s_0$. The rationality of F follows from the criterion of Borel since $N_{s,m} = 0$ for $s > s_0$.

We note that Peterson [13] also considered generalizations of Borel's theorem to $L\{t\}$ but his results correspond to the case in which S is the set of infinite primes of L and hence could not be used to exploit the results of p-adic analysis.

5. Applications. We note some immediate consequences of our main result.

1. If V is an affine or projective hypersurface and if α is an algebraic integer, $\alpha \neq 1$, such that the geometric mean of the ordinary absolute magnitudes of the conjugates of α over Q is less than q, then α^{-1} cannot be a zero (resp: pole) of $\zeta(V,t)$ if V is of even (resp: odd) dimension. Furthermore $t = 1$ is a pole of $\zeta(V,t)$ if V is a projective of even dimension and is not a pole if V is affine of odd dimension. In particular, if V is an irreducible, non-singular, projective variety of even dimension then the first Betti number ([2]) of V is zero. This is a direct consequence of [1], where we proved these statements in the projective case using the hypothesis of rationality. The affine statement is obtained by an obvious modification of this earlier treatment.

2. Each abstract variety, V, defined over k has a finite covering consisting of affine varieties defined over k whose intersections are affine. The rationality of the zeta function [12] of V now follows by an obvious combinatorial argument. In particular, the projective case could be treated in this way as a consequence of the affine case.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] B. Dwork, "On the congruence properties of the zeta function of algebraic varieties," *Journal für die reine und angewandte Mathematik*, to appear.

[2] A. Weil, "Solutions of equations in finite fields," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 497-508.

[3] L. Hua and H. Vandiver, "Characters over certain types of rings with application to the theory of equations in a finite field," *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 35 (1949), pp. 94-99.

[4] H. Davenport and H. Hasse, "Die Nullstellen der Kongruenzzetafunctionen in gewissen zylischen Fällen," *Journal für die reine und angewandte Mathematik*, vol. 172 (1948), pp. 151-182.

[5] E. Warning, "Bemerkung zur vorstehenden Arbeit von Herrn Chevalley," *Abh. Math. Sem. Hamburg*, vol. 11 (1935), pp. 76-83.

[6] B. Dwork, "Norm residue symbol in local number fields," *Abh. Math. Sem. Hamburg*, vol. 22 (1958), pp. 180-190.

[7] L. Stickelberger, "Über eine Verallgemeinerung der Kreistheilung," *Mathematische Annalen*, vol. 37 (1890), pp. 321-367.

[8] M. Krasner, "Theorie des corps values," *Comptes Rendus*, vol. 239, Pt. 1 (1954), pp. 745-747.

[9] E. Borel, "Sur une application d'un théorem de M. Hadamard," *Bull. Sci. Mathematiques*, Ser. 2, vol. 18 (1894), pp. 22-25.

[10] M. Krasner, "Theorie des functions," *Comptes Rendus*, vol. 222 (1946), pp. 37-40.

[11] L. Schnirelmann, "Sur les functions dan les corps normes et algebriquement fermes," *Bull. Acad Sci U. S. S. S. Math.*, vol. 5, pp. 487-497.

[12] S. Lang and A. Weil, "Number of points of varieties in finite fields," *American Journal of Mathematics*, vol. 76 (1954), pp. 819-827.

[13] H. Peterson, "Über Potenzreihen mit ganzen algebraischen Koeffizienten," *Abh. Math. Sem. Hamburg*, vol. 8 (1931), pp. 315-322.

WHITEHEAD PRODUCTS AND THE COHOMOLOGY STRUCTURE OF PRINCIPAL FIBRE SPACES.*

By Franklin P. Peterson.

1. **Introduction.** Let (E, p, B) be a principal fibre space with fibre F [3]. Let $\mu\colon F \times E \to E$ be the operation of the fibre on the total space. If (E, p, B) is part of the Postnikov system of a space X, then there is a connection between homotopy operations in X (e. g. Whitehead products), and $\mu^*\colon H^*(E) \to H^*(F \times E)$ (see [1]). In this paper we show how to compute μ^* in certain cases and give some applications to computing Whitehead products.

2. **Functional cohomology operations.** In this section we recall the properties of functional (primary) cohomology operations that we will need.

Let $\theta \in H^q(\pi, n; G)$, $f\colon X \to Y$, and $u \in H^n(Y; \pi)$. Assume that $f^*(u) = 0$ and $\theta(u) = 0$. Then one can define

$$\theta_f(u) \in H^{q-1}(X; G)/f^*(H^{q-1}(Y; G)) + {}^1\theta(H^{n-1}(X; \pi)),$$

where ${}^1\theta \in H^{q-1}(\pi, n-1; G)$ is the suspension of θ (see [3] for details). One of the main properties these operations have is the following one. If $g\colon W \to X$, then

$$g^*\theta_f(u) = \theta_{fg}(u) \in H^{q-1}(W; G)/g^*f^*(H^{q-1}(Y; G)) + {}^1\theta(H^{n-1}(W; \pi)).$$

Furthermore, one can define functional cup products and functional cohomology operations coming from sums of stable operations and cup products. For example, let $\theta \in H^q(\pi, n; G)$, $f\colon X \to Y$, $y \in H^{q-n}(Y; \pi')$, and $u \in H^n(Y; \pi)$. Assume that $f^*(u) = 0$ and $\theta(u) + y \smile u = 0$, where there is a given coefficient pairing $\pi' \otimes \pi \to G$. Then one can define

$$(\theta + y\smile)_f(u) \in H^{q-1}(X; G)/f^*(H^{q-1}(Y; G)) + ({}^1\theta + f^*(y)\smile)(H^{n-1}(X; \pi)).$$

3. **Principal fibre spaces.** In this section we study principal fibre spaces and prove our main theorem.

* Received June 18, 1959.

Let (E, p, B) be a principal fibre space with fibre F (see [3] for definition and elementary properties). Let $i\colon F \to E$ be the inclusion, $\nu\colon F \times F \to F$ the multiplication in F, $\mu\colon F \times E \to E$ the operation of F on E, and $\eta\colon F \times E \to E$ the projection on the second factor. The following diagrams are commutative.

$$
\begin{array}{ccc}
F \times F & \xrightarrow{\nu} & F \\
\Big\downarrow 1 \times i & & \Big\downarrow i \\
F \times E & \xrightarrow{\mu} & E
\end{array}
\qquad\qquad
\begin{array}{ccc}
F \times E & \xrightarrow{\mu} & E \\
\Big\downarrow \eta & & \Big\downarrow p \\
E & \xrightarrow{p} & B
\end{array}
$$

Let $f_0 \in F$ be the unit for ν. Then $i_E\colon E \to F \times E$ defined by $i_E(e) = (f_0, e)$ is such that $\mu i_E \simeq$ identity. Also, $i_F\colon F \to F \times E$ defined by $i_F(f) = (f, i(f_0))$ is such that $\mu i_F \simeq i$.

Let $\pi = \pi_n(F)$, and assume that F is $(n-1)$-connected. Let $\iota \in H^n(F; \pi) \approx \mathrm{Hom}(\pi, \pi)$ denote the canonical generator. Let $w \in H^{n+1}(B; \pi)$ be the image of ι under transgression (i.e., w is the characteristic class of (E, p, B)).

Let $k \in H^q(E; \pi')$. We wish to study $\mu^*(k)$. It is well-known that the exact cohomology sequence of $(F \times E, F \vee E)$ gives rise to the following exact sequence:

$$
0 \to H^q(F \times E, F \vee E; \pi') \xrightarrow{j^*} H^q(F \times E; \pi') \xrightarrow{h} H^q(F; \pi') + H^q(E; \pi') \to 0,
$$

where $h = i_F^* + i_E^*$. Thus

$$
h(\mu^*(k)) = i_F^*\mu^*(k) + i_E^*\mu^*(k) = i^*(k) + k.
$$

Hence we may write

$$
\mu^*(k) = 1 \otimes k + i^*(k) \otimes 1 + j^*(x),
$$

where $x \in H^q(F \times E, F \vee E; \pi')$.

Let $\theta \in H^{q+1}(\pi, n+1; \pi'), \psi \in H^{q-n}(B; G)$ be such that $\theta(w) + \psi \cup w = 0$, with a given coefficient pairing $G \otimes \pi \to \pi'$.

THEOREM 1. *Let k be a representative of*

$$
(\theta + \psi^{\cup})_p(w) \in H^q(E; \pi')/p^*(H^q(B; \pi')) + ({}^1\theta + p^*(\psi)^{\cup})(H^n(E; \pi)).
$$

Then

$$
\mu^*(k) = 1 \otimes k + {}^1\theta(\iota) \otimes 1 + (-1)^{(q-n)n} x' \otimes p^*(\psi),
$$

where ${}^1\theta(x') = {}^1\theta(\iota)$.

Proof.
$$\begin{aligned}\mu^*(k) &= \mu^*((\theta+\psi^\cup)_p(w)) = (\theta+\psi^\cup)_{p\mu}(w)\\ &= (\theta+\psi^\cup)_{p\eta}(w) - \eta^*((\theta+\psi^\cup)_p(w))\\ &= 1\otimes k \in H^q(F\times E;\pi')/\eta^*p^*(H^q(B;\pi'))\\ &\quad + ({}^1\theta+\eta^*p^*(\psi)^\cup)(H^n(F\times E;\pi)).\end{aligned}$$

Hence

$$\begin{aligned}\mu^*(k) &= 1\otimes k + 1\otimes u + [{}^1\theta + (1\otimes p^*(\psi))^\cup](x'\otimes 1 + 1\otimes x'')\\ &= 1\otimes k + 1\otimes u + 1\otimes {}^1\theta(x'') + 1\otimes(p^*(\psi)^\cup x'') + {}^1\theta(x')\otimes 1\\ &\quad + (-1)^{(q-n)n}x'\otimes p^*(\psi).\end{aligned}$$

Apply h and we see that $u + {}^1\theta(x'') + p^*(\psi)^\cup x'' = 0$, and ${}^1\theta(x') = i^*(k)$. By Lemma III. 3. 2 of [2], we have that $i^*(k) = {}^1\theta(\iota)$. Thus we have shown that ${}^1\theta(x') = {}^1\theta(\iota)$ and that

$$\mu^*(k) = 1\otimes k + {}^1\theta(\iota)\otimes 1 + (-1)^{(q-n)n}x'\otimes p^*(\psi).$$

4. Whitehead products. In this section, we show how Theorem 1 enables us to compute some Whitehead products in a space X from knowledge of the Postnikov system of X.

Let $p\colon X^{(n)}\to X^{(n-1)}$ be a fibre space with fibre $K(\pi,n)$ (e.g. part of the Postnikov system of a space X). Let $k\in H^q(X^{(n)};\pi_{q-1})$, $q>n+1$, and consider the fibre space $\bar p\colon X\to X^{(n)}$ with $K(\pi_{q-1},q-1)$ as fibre and k as k-invariant. As in Section 3, we may write $\mu^*(k) = 1\otimes k + i^*(k)\otimes 1 + j^*(x)$, where $x\in H^q(K(\pi,n)\times X^{(n)}, K(\pi,n)\vee X^{(n)};\pi_{q-1})$. Let

$$\alpha\in\pi_n(X)\approx\pi_n(K(\pi,n)),\qquad \beta\in\pi_{q-n}(X),\qquad \nu\colon\pi_i(X)\to H_i(X)$$

be the Hurewicz homomorphism, and let χ be the composition

$$\begin{aligned}&H^q(K(\pi,n)\times X^{(n)}, K(\pi,n)\vee X^{(n)};\pi_{q-1})\\ &\quad\to \operatorname{Hom}(H_q(K(\pi,n)\times X^{(n)}, K(\pi,n)\vee X^{(n)}),\pi_{q-1})\\ &\quad\to \operatorname{Hom}(H_n(K(\pi,n),\mathrm{pt.})\otimes H_{q-n}(X^{(n)},\mathrm{pt.}),\pi_{q-1}).\end{aligned}$$

Meyer [1] has proven the following theorem.

THEOREM 2. $[\alpha,\beta] = \chi(x)\ (\nu(\alpha)\otimes\bar p_*\nu(\beta))\in\pi_{q-1}(X)$.

Let $k^{n+1}\in H^{n+1}(X^{(n-1)};\pi)$ be the k-invariant for the fibre space $p\colon X^{(n)}\to X^{(n-1)}$. Let $\theta\in H^{q+1}(\pi,n+1;\pi')$, $\psi\in H^{q-n}(X^{(n-1)};G)$ be such that $\theta(k^{n+1}) + \psi^\cup k^{n+1} = 0$ with a given coefficient pairing $G\otimes\pi\to\pi_{q-1}$, and such that $k\in H^q(X^{(n)};\pi_{q-1})$ is a representative of $(\theta+\psi^\cup)_p(k^{n+1})$. [This con-

dition is many times fulfilled when $i^*(k) = {}^1\theta(\iota)$; e. g. see the example below.] We may then apply Theorems 1 and 2 to deduce the following theorem.

THEOREM 3. *$[\alpha,\beta] = (-1)^{(q-n)n}x'(\nu(\alpha))\cdot\psi(p_*\bar{p}_*\nu(\beta)) \in \pi_{q-1}(X)$, where the pairing is the given coefficient pairing $\pi\otimes G\to\pi_{q-1}$, and where ${}^1\theta(x') = {}^1\theta(\iota)$.*

We give the following example to illustrate the applications of Theorem 3. The result here is known (see [4]); however, our method has the advantage of being purely cohomological in nature. Also, the method gives further insight into the structure of the spectral sequence of a fibre space.

Let CP^m be complex projective m-space. We shall calculate the Whitehead product pairing

$$\pi_{2m+1}(CP^m)\otimes\pi_2(CP^m)\to\pi_{2m+2}(CP^m).$$

(Recall that $\pi_{2m+1}(CP^m) = Z = \pi_2(CP^m)$ and $\pi_{2m+2}(CP^m) = Z_2$.) Let $\bar{\iota}\in H^2(Z,2;Z)$, then $\bar{\iota}^{m+1}\in H^{2m+2}(Z,2;Z)$ is the first k-invariant. To study the above Whitehead product in CP^m, it is enough to study it in X, the Postnikov system of CP^m through dimension $2m+2$. I. e., $\bar{p}\colon X\to X^{(2m+1)}$ and $p\colon X^{(2m+1)}\to K(Z,2)$, the first with fibre $K(Z_2,2m+2)$ and k-invariant the non-zero class in $H^{2m+3}(X^{(2m+1)};Z_2) = Z_2$, where $i^*(k) = \mathrm{Sq}^2(\iota)$, and the second with fibre $K(Z,2m+1)$ and k-invariant $\bar{\iota}^{m+1}$. (For details of this computation, see [2].) In case m is odd, $\mathrm{Sq}^2(\bar{\iota}^{m+1}) = 0$. In case m is even, $\mathrm{Sq}^2(\bar{\iota}^{m+1}) + \bar{\iota}\cup\iota^{m+1} = 0 \bmod 2$, with the non-zero pairing $Z\otimes Z\to Z_2$. Thus we may apply Theorem 3 with $\psi = 0$ or $\bar{\iota}$ respectively. Since $\mathrm{Sq}^2(x') = \mathrm{Sq}^2(\iota)\neq 0$ in $H^{2m+3}(Z,2m+1;Z_2)$, x' is an odd multiple of ι. If α is a generator of $\pi_{2m+1}(X) = \pi_{2m+1}(CP^m)$, and β a generator of $\pi_2(X)$, then $[\alpha,\beta] = 0$, if m is odd and, if m is even, $[\alpha,\beta] = \iota(\nu(\alpha))\cdot\bar{\iota}(p_*\bar{p}_*\nu(\beta))$ = non-zero element of $\pi_{2m+2}(X) = Z_2$ as $\nu(\alpha)$ is dual to ι and $p_*\bar{p}_*\nu(\beta)$ is dual to $\bar{\iota}$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] J.-P. Meyer, "Whitehead Products and Postnikov Systems," *American Journal of Mathematics*, ——————.

[2] F. P. Peterson and N. Stein, "Two Formulas Concerning Secondary Cohomology Operations," *ibid.*, vol. 81 (1959), pp. 281-305.

[3] ——— and E. Thomas, "A Note on Non-stable Cohomology Operations," *Boletín de la Sociedad Matemática Mexicana*, vol. 3 (1958), pp. 13-18.

[4] N. Stein, Cornell thesis, 1957.

SOLVABLE LIE GROUPS ACTING ON NILMANIFOLDS.*

By Louis Auslander.[1]

Introduction. In [1] Maurice Auslander and the author classified all the connected, simply connected solvable Lie groups which can operate transitively and with discrete isotropy group on the n-dimensional torus. If Z^n denotes the additive group of integers taken n times, the above classification is equivalent to the classification of all the connected, simply connected solvable Lie groups S which can contain Z^n as discrete uniform subgroup. If R^s denotes the direct sum of the additive group of reals taken s times, our result can be stated as follows: A necessary and sufficient condition for a connected, simply connected solvable Lie group S to contain Z^n as a discrete uniform subgroup is that S satisfy the diagram

$$1 \to R^s \to S \to R^t \to 1,$$

where $s + t = n$, the automorphisms of R^s induced by R^t form a compact group and the extension is the *split* extension. If T denotes the torus group of automorphisms of the group R^sZ^n, then T is uniquely extendable to a torus group of automorphisms of R^n and S is easily seen to be a subgroup of $R^n \cdot T$, where the dot denotes the semi-direct product.

This paper has as its goal the extension of the above results to the case where Z^n is replaced by the fundamental group of any compact nilmanifold. We will always use Γ to denote such a group. Or restating our goals, we will study the connected, simply connected, solvable Lie groups S which can contain the group Γ as a discrete uniform subgroup. By a result of Malcev [3] there exists a *unique* connected, simply connected nilpotent group $N(\Gamma)$ which contains Γ as discrete uniform subgroup. We will first show that for a connected, simply connected solvable Lie group S containing Γ as discrete uniform subgroup there exists a torus group T of automorphisms of $L(\Gamma)$ such that $S \subset N(\Gamma) \cdot T$, where the dot again denotes the semi-direct product. This will be the easier part of our program. The second part will be devoted

* Revised October 1, 1959.

[1] During part of the time this paper was being written the author was supported by a National Science Foundation Grant.

to showing that the group extension is essentially controlled by the discrete uniform subgroup. The explicit form of our result is given in Theorem 2.

Once we have proven these results for the fundamental groups of nil-manifolds it is reasonable to try and extend these results to the fundamental groups of compact solvmanifolds. We did not succeed in this attempt. This may perhaps best be explained by remarking that our whole proof of Theorem 2 rests on a series of steps which enable us to reduce a non-abelian group extension problem to the abelian group extension problem which had already been solved in [1]. Hence, the work in [1], rather than being replaced by this paper, is made to form the basis of the results here obtained. However, for the general case of solvmanifolds, our reduction fails.

For the rest of this paper we will use the expression solvable Lie group to denote connected, simply connected solvable Lie groups.

1. Theorem 1.

THEOREM 1. *Let S be a solvable Lie group and contain Γ as discrete uniform subgroup. Further, let Γ be a discrete uniform subgroup of the nil-potent group N. Then there exists a torus group T of automorphisms of N such that $S \subset N \cdot T$, where the dot denotes the semi-direct product.*

Proof.[2] Let H be the maximal normal analytic nilpotent subgroup of S. Then ΓH is normal in S and by the results in [3, 4] can be identified with a normal subgroup of N, for it contains the commutator subgroup of both N and S.

The inner automorphisms of S induce automorphisms of ΓH, and hence isomorphisms of Γ into N, and hence by [3] isomorphisms of N onto N. Thus S operates on N. We now form the semidirect product $N \cdot S$ and let Δ denote the subgroup of elements (x, x^{-1}) for $x \in \Gamma H$. Then Δ is normal in $N \cdot S$ and N and S have isomorphic images in $N \cdot S/\Delta$ under the natural projection. Further, the image of N is a normal subgroup and the factor group of $N \cdot S/\Delta$ by the image of N is isomorphic to the toroid group $S/\Gamma H$. Since any extension of a connected, simply connected solvable group by a compact group is a split extension, we have $N \cdot S/\Delta \approx N \cdot T$, with T a torus group. Hence S has an isomorphic image in $N \cdot T$.

2. Statement of Theorem 2. Let G and H be connected Lie groups. An extension of G by H is a pair (E, π) consisting of a connected Lie group

[2] The proof of this theorem was suggested to the author by the referee and is a considerable improvement of the author's original proof.

E which contains G as a closed normal subgroup and a continuous homomorphism of E onto H whose kernel is G. Let $A(G)$ denote all continuous automorphisms of G and let $I(G)$ denote the normal subgroup of $A(G)$ consisting of inner automorphisms. Then if (E, π) is an extension of G by H, this determines a homomorphism ϕ of H into $A(G)/I(G)$. We will call this homomorphism the *action* of H on G. We will also say that the group extension (E, π) is associated with the triple (G, H, ϕ) and vice versa. Clearly there may be many or no group extensions associated with a given triple.

Let Γ be a non-abelian uniform subgroup of a connected, simply connected, nilpotent Lie group N. Let S be any connected, simply connected, solvable Lie group which contains Γ as a discrete uniform subgroup. Let H be the maximal normal analytic nilpotent subgroup of S. Then H contains the commutator subgroup of S and is a closed normal subgroup of S. Let π: $S \to S/H$ and let $V = S/H$. Then V is a connected, simply connected, abelian Lie group and V contains $\Gamma H/H$ as a discrete uniform subgroup. Hence associated with the group extension (S, π) is the triple (H, V, ϕ), where ϕ is the action of V on H.

THEOREM 2. *Let S_1 and S_2 be connected, simply connected, solvable Lie groups with discrete uniform subgroups Γ_1 and Γ_2 and let Γ_1 and Γ_2 be fundamental groups of compact nilmanifolds. Assume that α is an isomorphism of Γ_1 onto Γ_2 with the following properties:*

1. *If H_1 and H_2 are the maximal normal analytic nilpotent subgroups of S_1 and S_2 respectively, then α induces an isomorphism α_1 of $\Gamma_1 \cap H_1$ onto $\Gamma_2 \cap H_2$. By the results in* [3], *we see that α is uniquely extendable to an isomorphism α^*_1 of H_1 onto H_2.*

2. *Let α induce the isomorphism α_2 of $\Gamma_1/\Gamma_1 \cap H_1$ onto $\Gamma_2/\Gamma_2 \cap H_2$. Then α_2 can be extended to give a unique isomorphism α^*_2 of V_1 onto V_2. Further, if S_1 and S_2 are associated with the triples (H_1, V_1, ϕ_1) and (H_2, V_2, ϕ_2) respectively, we shall assume that the following diagram is commutative.*

$$\begin{tikzcd} V_1 \arrow[r, "\phi_1"] \arrow[d, "\alpha^*_2"'] & A(H_1)/I(H_1) \arrow[d, "\alpha^{**}_1"] \\ V_2 \arrow[r, "\phi_2"] & A(H_2)/I(H_2), \end{tikzcd}$$

where α^*_1 is defined in the obvious way from α^*_1.

Under all the above assumptions S_1 is isomorphic to S_2.

3. Preliminaries to Theorem *C*. Let S contain Γ as a discrete uniform subgroup, where Γ is a discrete uniform subgroup of the nilpotent Lie group N. Let H be the maximal normal analytic nilpotent subgroup of S and let $H^* = H\Gamma$. Then H^* is a nilpotent group and we will denote by $H^*{}_j$ the j entry in the lower central series of H^*. Let $C(H^*{}_j)$ denote the closure of $H^*{}_j$. Let r be the largest index such that $C(H^*{}_r)$ has a non-trivial identity component. Let H_r denote the identity component of $C(H^*{}_r)$.

LEMMA 1. *H_r is in the center of H^* and H_r is a normal analytic subgroup of S.*

Proof. Let $h \in H^*$ and let D be the subset of H^* consisting of all elements $hxh^{-1}x^{-1}$ for $x \in H_r$. Clearly D is a connected set and contains the identity element. We assert that D is a subset of $C(H^*{}_{r+1})$. To see this, notice that the mapping $\phi(x) = hxh^{-1}x^{-1}$ is a continuous mapping of H^* into itself. Hence for any subset A of H^*

$$\phi(C(A)) \subset C(\phi(A)),$$

where $C(\ )$ denotes the closure of the set in the parenthesis. Hence

$$\phi(C(H^*{}_r)) \subset C(\phi(H^*{}_r)) \subset C(H^*{}_{r+1}).$$

But $C(H^*{}_r) \supset H_r$. This proves the assertion that $D \subset C(H^*{}_{r+1})$. Hence, since the component of the identity is trival in $C(H^*{}_{r+1})$ and D is connected and contains the identity, we see that D must consist of the identity element. Hence H is in the center of H^*.

It remains to show that H_r is a normal subgroup of S. Now H^* is a normal subgroup of S and $H^*{}_r$ is a characteristic subgroup of H^*, i.e., every automorphism of H^* leaves $H^*{}_r$ invariant. Hence every continuous automorphism of H^* leaves $C(H^*{}_r)$ invariant. Hence $C(H^*{}_r)$ is normal in S. From this it follows that the identity component of $C(H^*{}_r)$, H_r, is normal in S.

LEMMA 2. *$H_r \cap \Gamma$ is a discrete uniform subgroup in H_r.*

Proof. $\Gamma_0 = H \cap \Gamma$ is a discrete uniform subgroup of H. But Γ_0 is a subgroup of N also. This induces an isomorphism of H into N. We will call this image of H in N also. Consider $C(H^*{}_r) \subset N$. Then $H^*{}_r \subset N_r$, where N_r is the r entry in the lower central series of N, and, since N_r is closed, $C(H^*{}_r) \subset N_r$ also. Let Γ_r be the r entry in the lower central series in Γ. Then $\Gamma_r \subset H^*{}_r$ and hence in $C(H^*{}_r)$. Now Γ_r is normal in N_r (see [3]) and hence in $C(H^*{}_r)$. Since $C(H^*{}_r)$ is locally connected, H_r is open and

closed in $C(H^*_r)$. Thus $\Gamma_r H_r$ is closed in $C(H^*_r)$ and hence in N_r. It follows that Γ_r is uniform in $\Gamma_r H_r$, that is $\Gamma_r H_r/\Gamma_r$ is compact. Hence $\Gamma_r \cap H_r$ is uniform in H_r.

COROLLARY 1. *S/H_r has a discrete uniform subgroup isomorphic to $\Gamma H_r/H_r$. Further, H/H_r has a discrete uniform subgroup isomorphic to $\Gamma/H_r\Gamma$.*

The proof of this corollary and Lemma 3 below follow easily from Lemma 2.

LEMMA 3. *Let S_1, S_2 and Γ_1, Γ_2 satisfy the hypothesis of Theorem 2. Then*

$$S_1/(H_1)_r, \qquad S_2/(H_2)_r$$

and

$$\Gamma_1/(H_1)_r \cap \Gamma_1 \qquad \Gamma_2/(H_2)_r \cap \Gamma_2$$

satisfy the hypothesis to Theorem 2.

LEMMA 4. *Let S contain Γ as a discrete uniform subgroup. Then S contains the commutator subgroup of N as a normal subgroup.*

Proof. The statement of the lemma is meaningful as we have identified H in S with H in N. Further, H contains the commutator of N and hence by the identification of H in S with H in N, this gives rise to a group in S. The proof of this lemma will be induction on the dimension of S. Clearly the lemma is true if $\dim. S = 1$. Let us assume now that the lemma is true for $\dim. S < n$. Assume now that $\dim. S = n$. Then consider $H_r \subset S$ and the exact sequence

$$1 \to H_r \xrightarrow{i} S \xrightarrow{j} S/H_r \to 1.$$

Then by the previous discussion we see that S/H_r satisfies the induction hypothesis and hence the commutator subgroup C^* of N/N_r is a normal subgroup of S/H_r. Hence $j^{-1}(C^*) = C$ is a normal subgroup of S. But considering $H \subset S$, we get the diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & H_r & \longrightarrow & H & \longrightarrow & H/H_r & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow{\scriptstyle k} & & \downarrow & & \\
1 & \longrightarrow & H_r & \longrightarrow & N & \xrightarrow{j^*} & N/H_r & \longrightarrow & 1,
\end{array}$$

where $k^*(C^*)$ is the commutator subgroup of N/H_r and $j^{*-1}(k^*(C^*)) = k(C)$. Since H_r is contained in the commutator subgroup of N, $k(C)$ is the commutator subgroup of N. This proves the lemma.

4. Theorem 2.

Proof of Theorem 2. It is clear from the hypothesis that dim. S_1 $=$ dim. S_2. We will prove the Theorem by induction on the dimension of S_1. The theorem is trivially true if dim. $S_1 = 1$. Assume the theorem is true if dim. $S_1 < n$, and assume that the dim. $S_1 = n$. In order to simplify the presentation and notation, we will use the isomorphisms stated in the hypothesis to this theorem to identify objects; i. e. we will use H to denote H_1 and H_2, H_r to denote $(H_1)_r$, $(H_2)_r$ and Γ to denote Γ_1 and Γ_2, etc. Whenever this is done we will explicitly be using the isomorphisms in the hypothesis to this theorem. We may then form the groups S_1/H_r and S_2/H_r. As a consequence of Lemma 3 and our induction hypothesis we have that S_1/H_r and S_2/H_r are isomorphic. We will identify them under this isomorphism and denote them by S^*. Hence we have

$$1 \to H_r \to S_i \to S^* \to 1, \qquad i = 1, 2,$$

where H_r is abelian.

LEMMA 5. *We can choose cocycles f_1 and f_2 representing the group extensions S_1 and S_2 respectively such that:*

a) *$f_1 - f_2 = 0$ when restricted to $\Gamma H/H_r \subset S^*$*

b) *Corresponding to $f_1 - f_2 = f$ there is a* Lie group *satisfying the diagram:*

$$1 \to H_r \to S' \to S^* \to 1$$

with S^1 having f as extension cocycle.

Proof. In order to keep the exposition at a minimum we will assume in presenting the details of this proof that the reader is familiar with the ideas in Hochschild's paper [2]. In particular, we will prove the lemma by constructing two continuous cocyles or factor sets f_1 and f_2 such that $f_1 - f_2 = 0$ when restricted to $\Gamma H/H_r$.

Consider $H\Gamma \subset S_i$ and $H\Gamma/H_r \subset S^*$. Now a factor set may be generated by a cross section in the bundle S_i over the base space S^* with fiber H_r. Further, if the cross sections are continuous, then the factor sets corresponding to them will be continuous. Now for S_i, $i = 1, 2$, choose continuous sections $\rho^*_1 = \rho^*_2$ over $H\Gamma/H_r \subset S^*$ with images in $H\Gamma$. Then since the fiber H_r is homeomorphic to euclidean space, each of the two partial sections can be extended to continuous sections ρ_1 and ρ_2 respectively. Let f_i be the factor set associated with the section ρ_i. Then clearly f_i is continuous and $f_1 - f_2 = 0$

when restricted to $H\Gamma/H_r$ since $\rho_1 = \rho_2$ when restricted to $H\Gamma/H_r$. This proves the lemma.

Since H_r is in the center of ΓH, $H/H_r = S^*$ acts trivially on H. Further, f restricted to $\Gamma H/H_r$ is cohomologous to zero. From these considerations, we see that S' has a discrete uniform subgroup Γ' such that Γ' is also a discrete uniform subgroup of the nilpotent Lie group $N' = H_r \oplus N/H_r$, where $\oplus$ denotes the direct product of groups.

Let G be the pre-image of H/H_r in S'. Then H_r is a normal subgroup in G.

LEMMA 6. *$G = H_r \oplus K$, where K is also a normal subgroup in S' and S'/K has a free abelian discrete uniform subgroup.*

(We will now complete the proof of Theorem C under the assumption that Lemma 6 is true. We will present a proof of Lemma 6 at the end of the demonstration of Theorem 2.)

By Lemma 6, we have that K is a normal subgroup of S' and K is a pre-image of H/H_r in S'. Hence we may form

$$1 \to H_r \to S'/K \to V \to 1,$$

where S'/K has a discrete abelian uniform subgroup. Therefore, by the results in [1], S'/K is a split extension.

Now consider the diagram

$$\begin{array}{ccccccccc} 1 & \to & H_r & \to & S' & \to & S^* & \to & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \to & H_r & \to & S'/K & \to & S^*/K & \to & 1. \end{array}$$

This shows that S' is also a split extension and hence $f_1 - f_2$ is cohomologous to zero. Hence S_1 is isomorphic to S_2.

Proof of Lemma 6. By Lemma 4, the commutator subgroup C of N' is a normal subgroup of S' and $C \cap \Gamma'$ is a discrete uniform subgroup of C. Now $C \subset G$ and intersects H_r in the identity element. Hence we may form

$$1 \to G/C \to S'/C \to V \to 1.$$

Now G/C is abelian and the image of Γ' in S'/C is a discrete uniform subgroup. Further, the image of Γ' in V is a discrete uniform subgroup which acts trivially on G/C. If $i(H_r)$ denotes the image of H in G/C, there is a normal subgroup K^* in G/C intersecting $i(H_r)$ in the identity and invariant

under V acting on G/C. This follows from the fact that V acts as a compact group on G/C and $i(H_r)$ is an invariant space of V acting on G/C. Hence K^* is normal in S'/C. Let K be the pre-image in S' of K^* in S'/C. Clearly K satisfies the conclusions of the Lemma provided S'/C has a discrete abelian uniform subgroup, it is a split extension, i. e. there exists V^*, a subgroup of S'/C, which is mapped isomorphically onto V by the projection mapping. Further let Γ^* be any pre-image of Γ' in V^*. Then Γ^* acts trivially on $i(H_r)$ in G/C.

Choose a discrete uniform subgroup of the image of $i(H_r)$ in S'/K, and consider the image of Γ^* in S'/K. Then these groups generate a discrete uniform abelian subgroup in S'/K.

REFERENCES.

[1] L. Auslander and M. Auslander, "Solvable Lie groups and locally Euclidean Riemann spaces," to appear in the *Proceedings of the American Mathematical Society.*

[2] G. Hochschild, "Group extensions of Lie groups," *Annals of Mathematics,* vol. 54 (1951), pp. 96-109.

[3] A. Malcev, "On a class of homogeneous spaces," *American Mathematical Society Translations,* No. 39.

[4] G. D. Mostow, "Factor spaces of solvable groups," *Annals of Mathematics,* vol. 60 (1954), pp. 1-27.

THE MANIFOLDS COVERED BY A RIEMANNIAN HOMOGENEOUS MANIFOLD.*

By JOSEPH A. WOLF.[1]

Introduction. The sphere is known to be the universal covering for complete connected Riemannian manifolds of constant positive curvature. More precisely, if M is an n-dimensional complete connected Riemannian manifold of constant sectional curvature $k^2 > 0$ with $k > 0$, and if S^n is the sphere of radius k^{-1} in Euclidean space R^{n+1}, with the induced metric, then there is a covering of M by S^n such that the covering projection is a local isometry. Because of this phenonmenon, the complete connected Riemannian manifolds of constant positive curvature are called the "spherical space-forms." In his thesis, G. Vincent [14] attempted to classify them. Following this line of investigation, we take a compact connected Riemannian homogeneous manifold M and ask which Riemannian manifolds admit M as a Riemannian covering manifold. In Chapter I, this problem is reduced to a problem on discrete subgroups of compact Lie groups:

Given a compact Lie group G and a closed subgroup K, find all finite subgroups Γ of G such that Γ meets the union of the conjugates of K only at the identity element of G.

For the most part we restrict our attention to the case where Γ lies in the identity component of G, or, equivalently, where G is connected. In Chapter II we obtain some bounds on the ranks of abelian subgroups of Γ, and see that the problem of classifying these groups Γ is inaccessible unless rank. G — rank. $K \leqq 1$.

Chapter II ends with a sharper bound on the ranks of abelian subgroups of Γ, in case rank. G — rank. $K = 1$, which implies that every abelian subgroup of Γ is cyclic if the semisimple part of G is simply connected and Γ lies in the identity component of G. We remark that H. Zassenhaus [16] and M. Suzuki [13] have given a complete classification of the finite groups

* Received August 25, 1959; revised May 23, 1960.

[1] This work is part of the author's doctoral dissertation, and was supported by National Science Foundation predoctoral fellowships.

with all abelian subgroups cyclic. Under certain conditions on G, K and the order of Γ (Corollary 10.1) it follows that Γ itself is cyclic.

In Chapter III we obtain an arithmetic criterion (Theorem 6), assuming G and K connected, for an arbitrary given finite cyclic subgroup of G to act freely on G/K. This criterion involves the Weyl group of G and the position of K in G. It is applied to an arbitrary finite subgroup Σ of G by finding cyclic subgroups of Σ such that every element of Σ is $\mathrm{ad}(G)$-conjugate to an element of one of these cyclic subgroups. Applying to the case where G is a classical group we obtain some information on elements of order 2 in Γ, assuming $\mathrm{rank.}\, G - \mathrm{rank.}\, K = 1$. Finally, we sharpen the bound on the ranks of abelian subgroups of Γ in case G is a special orthogonal group. In Chapter V we apply the arithmetic criterion again to the case where G is a classical group. Chapter VI is a consideration of the case where $\mathrm{rank.}\, G = \mathrm{rank.}\, K$, i.e., the Euler characteristic $\chi(M) \neq 0$, and shows that M covers only a finite number, up to isometry, of Riemannian manifolds.

Our problem can be considered as a generalization of the classical Clifford-Klein problem of finding all spherical space-forms, in that we have replaced the sphere by an arbitrary (for Theorems 1 to 4 and Theorem 6), or at least more general, compact connected Riemannian homogeneous manifold. Another direction of generalization is that of considering finite groups which admit a free topological action on a space similar in some way to a sphere. In this regard, we mention some of the work of P. A. Smith [11], P. E. Conner [5], J. Milnor [8] and A. Heller [6].

I especially wish to thank Professor S. S. Chern, under whose guidance this paper was written, for many helpful discussion and comments. I also wish to thank Professors A. Borel, H. C. Wang and R. S. Palais for many helpful discussions. Some of Professor Borel's work [1, 2] is crucial to this paper, and Professor Palais pointed out a lemma of Mostow used in the proof of Theorem 11.

Chapter I. Reduction to a problem on Lie groups.

I. 1. Covering spaces. In order to establish notation and terminology we will recall some well-known facts and definitions concerning covering spaces. All spaces will be Hausdorff and all maps will be continuous.

A *covering* is a map $p: X \to X'$ of arcwise connected, locally simply connected spaces where every element of X' has a neighborhood U such that p maps each component of $p^{-1}(U)$ homeomorphically onto U. $p^{-1}(x')$ is the *fibre* over $x' \in X'$. All fibres have the same cardinality, the *multiplicity* of

the covering. A *finite covering* is a covering of finite multiplicity. p induces a monomorphism of fundamental groups, and the covering is *normal* if $p\cdot\pi_1(X)$ is a normal subgroup of $\pi_1(X')$. This is independent of choice of basepoints. If H is a subgroup of $\pi_1(X')$, there is a covering $q\colon Y\to X'$ and a choice of basepoint in Y such that $q\cdot\pi_1(Y)=H$.

THEOREM 1. *If $p\colon X\to X'$ is a finite covering, there is a finite normal covering $q\colon X''\to X$, where $pq\colon X''\to X'$ is a finite normal covering.*

Proof. The multiplicity of p being equal to the index of $H=p\cdot\pi_1(X)$ in $\pi_1(X')$, the normalizer N of H in $\pi_1(X')$ has finite index in $\pi_1(X')$. Consequently ([7], p. 82) H has only a finite number of conjugates in $\pi_1(X')$, so ([7], p. 62) the intersection J of the conjugates of H has finite index in $\pi_1(X')$. Let $q\colon X''\to X$ be a covering, where $q\cdot\pi_1(X'')=p^{-1}(J)$, and the normality conditions follows from the construction of J. *QED.*

An action of a group Γ on a space X is *effective* if the identity element of Γ is the only element inducing the identity transformation of X, is *free* if the identity element of Γ is the only element which leaves fixed a point of X, and is *properly discontinuous* if every point of X has a neighborhood which does not meet any of its transforms under Γ. If X is compact, the action of Γ is properly discontinuous if and only if Γ is finite and acts freely. The set $\Gamma(x)$, the *orbit* of a point x of X, is the set of images of x under Γ. The space X/Γ of orbits is given the quotient topology for the natural projection $x\to\Gamma(x)$; the natural projection $X\to X/\Gamma$ is a covering if and only if Γ acts properly discontinuously on X.

A *deck transformation* of a covering $p\colon X\to X'$ is a homeomorphism $\gamma\colon X\to X$, where $p\cdot\gamma=p$. The group Γ of all deck transformations acts properly discontinuously on X, and ([12], § 14) p is a normal covering if and only if Γ is simply transitive on each fibre, i.e., if and only if $p\colon X\to X'$ is a principal bundle with group Γ, i.e., if and only if $X'=X/\Gamma$.

A *Riemannian covering* is a covering $p\colon M\to M'$, where M and M' are Riemannian manifolds and p is a local isometry. If just one of M and M' is a Riemannian manifold, the requirement that p be a local isometry gives a Riemannian structure to the other and makes p a Riemannian covering. We easily see that a deck transformation is an isometry of M, because p is a local isometry.

A *Riemannian homogeneous manifold* is a Riemannian manifold whose group of isometries is transitive.

THEOREM 2. *If $q: M'' \to M$ is a Riemannian covering and M is Riemannian homogeneous, then M'' is Riemannian homogeneous.*

Proof. A one-parameter group of isometries of M is a homotopy and can be lifted to M'' by the covering homotopy theorem. The lifted homotopy consists of isometries of M'' because q is a local isometry. It follows that the group of isometries of M'' is locally transitive and therefore transitive. *QED.*

I. 2. Reduction to a problem on discrete subgroups of compact Lie groups.

THEOREM 3. *Let M be a compact connected Riemannian homogeneous manifold, G the group of isometries of M, K an isotropy subgroup of G, and Γ a subgroup of G. Then Γ is a properly discontinuous group of isometries of M if and only if Γ is finite and $\Gamma \cap \mathrm{ad}(G)K = 1$, where 1 is the identity element of G and $\mathrm{ad}(G)K$ is the set of all $\mathrm{ad}(g)k = gkg^{-1}$ with $g \in G$ and $k \in K$.*

Proof. M is compact and Γ is a group of isometries of M, so Γ is a properly discontinuous group of isometries of M if and only if Γ is finite and acts freely on M. G is transitive on M, so the isotropy subgroups of G are the subgroups $\mathrm{ad}(g)K$ with $g \in G$. Γ acts freely on M if and only if it meets each isotropy subgroup only at 1. Hence Γ acts freely if and only if $\Gamma \cap \mathrm{ad}(G)K = 1$. *QED.*

Using Theorems 1, 2 and 3, we see that the original problem

Given a compact connected Riemannian homogeneous manifold M, find all Riemannian manifolds which admit a Riemannian covering by M.

is reduced to the problem

Given a compact Lie group G and a closed subgroup K, find all finite subgroups Γ of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$.

by taking G to be the group of isometries of a finite Riemannian covering manifold of M and K to be an isotropy subgroup of G. We then note that G and K are both compact, each has only a finite number of components, and K meets every component of G.

Chapter II. Necessary conditions for a finite group to act as a properly discontinuous group of isometries of a compact connected Riemannian homogeneous manifold.

Given a compact Lie group G and a closed subgroup K, we will find some necessary conditions on finite subgroups Γ of G for $\Gamma \cap \mathrm{ad}(G)K = 1$. These will involve the ranks of G, K, and some subgroups of Γ.

The *rank* of a finite abelian group B is the minimal number of factors in a direct product decomposition of B into cyclic groups, and is denoted rank. B. For example, if p is a prime, the elementary p-group with p^h elements, $(Z_p)^h = Z_p \times \cdots \times Z_p$, the direct product of h copies of the cyclic group of order p, has rank h. The rank of B is the maximum of the ranks of its elementary p-subgroups, and B is cyclic if and only if rank. $B \leqq 1$. If p is a prime, the *p-rank* of B, denoted p-rank. B, is the rank of a p-Sylow subgroup of B. It is the maximal integer h such that B has a subgroup isomorphic to $(Z_p)^h$.

The *rank* of a compact Lie group H, denoted rank. H, is, as usual, the common dimension of the maximal toral subgroups of H.

II. 1. A bound on the ranks of certain abelian subgroups.

THEOREM 4. *Let K be a closed subgroup of a compact Lie group G.*

1. *Given a finite subgroup Γ of G such that $\Gamma \cap \mathrm{ad}(G) = 1$ and an abelian subgroup B of Γ which lies in a torus of G, we have* rank. $B \leqq$ rank. G — rank. K.

2. *The above bound is the best possible in the sense that there is a positive integer $m(G,K)$ such that, given a finite abelian group A with* rank. $A \leqq$ rank. G — rank. K *and $m(G,K)$ prime to the order of A, a torus of G has a subgroup A' which is isomorphic to A and such that $A' \cap \mathrm{ad}(G)K = 1$.*

Proof. Let T' be a maximal torus of K, T a maximal torus of G which contains T', $n =$ rank. G and $k =$ rank. K. We replace B by a conjugate which lies in T and still have $B \cap \mathrm{ad}(G)K = 1$, hence $B \cap T' = 1$. It follows that the canonical map of T onto the $(n-k)$-torus T/T' maps B monomorphically. Since a finite subgroup of an $(n-k)$-torus has rank at most $n-k$, we conclude rank. $B \leqq n-k$. This proves the first statement.

Let K_0 be the identity component of K, G_0 the identity component of G, $W = \{w_1, \cdots, w_q\}$ an enumeration of the Weyl group of G_0 with respect

to T, and $\{a_1, \cdots, a_t\}$ a set of automorphisms of G_0 which preserve T such that the automorphism group $\mathrm{ad}(G)$ of G_0 can be written as the union of the $a_i \cdot \mathrm{ad}(G_0)$. Two elements of T are $\mathrm{ad}(G_0)$-conjugate if and only if they are W-conjugate, and it follows that an element of T lies in $\mathrm{ad}(G)K_0$ if and only if it lies in one of the $T_{ij} = a_i(w_j(T'))$. As there are only a finite number of the k-tori T_{ij}, there exists an $(n-k)$-torus V in T which intersects each T_{ij} in a finite group. Let $m(G,K)$ be the product of the primes occurring in the orders of these finite groups and in the order b of K/K_0. Let $\beta \in V$ have order prime to $m(G,K)$ and lie in $\mathrm{ad}(G)K$. Then $\beta^b \in \mathrm{Ad}(G)K_0$, so $\beta^b \in T_{ij}$ for some (i,j). Since the order of β^b is also prime to $m(G,K)$, this implies, by the definition of $m(G,K)$, that $\beta^b = 1$. Since the order of β is prime to b, this implies $\beta = 1$.

We can find a subgroup A' of V which is isomorphic to A because V is an $(n-k)$-torus and A is a finite abelian group of rank at most $n-k$. The considerations above show that $A' \cap \mathrm{ad}(G)K = 1$ if the order of A, hence of A', is prime to $m(G,K)$. *QED.*

In Chapter III we will see examples where K, and even G, is connected and $m(G,K)$ must be even, hence $m(G,K) > 1$.

II. 2. The work of A. Borel on torsion and subgroups which lie in a torus. A. Borel has proved ([1], Chapter XII) that if G is a compact connected Lie group with classifying space B_G, p is a prime, and the integral cohomology ring $\boldsymbol{H}^*(B_G, \boldsymbol{Z})$ has no p-torsion, then every elementary p-subgroup (subgroup isomorphic to some $(Z_p)^k$) of G lies in a torus of G. A case by case check proves the converse. Borel has also shown that $\boldsymbol{H}^*(G, \boldsymbol{Z})$ has p-torsion if and only if $\boldsymbol{H}^*(B_G, \boldsymbol{Z})$ has p-torsion, using known results and checking the case $p = 5$ for $\boldsymbol{E}_6$, $p = 5$ and $p = 7$ for $\boldsymbol{E}_7$ and $p = 7$ for $\boldsymbol{E}_8$. The summary of the situation is that the following are equivalent:

1. $\boldsymbol{H}^*(G, \boldsymbol{Z})$ has no p-torsion.

2. $\boldsymbol{H}^*(B_G, \boldsymbol{Z})$ has no p-torsion.

3. G_{ss} being the semisimple part of G, $\boldsymbol{H}^*(G_{ss}, \boldsymbol{Z})$ has no p-torsion.

4. $\pi_1(G_{ss})$ has order prime to p, and, if G' is a simple factor of the universal covering group of G_{ss}, then $\boldsymbol{H}^*(G', \boldsymbol{Z})$ has no p-torsion.

Finally, if H is a compact, connected, simple, simply connected Lie group, then $\boldsymbol{H}^*(H, \boldsymbol{Z})$ has p-torsion in precisely these cases:

1. $p = 2$ and $H = \boldsymbol{E}_8$, $\boldsymbol{E}_7$, $\boldsymbol{E}_6$, $\boldsymbol{F}_4$, $\boldsymbol{G}_2$, or $\boldsymbol{Spin}(n)$ with $n \geqq 7$.

2. $p=3$ and $H=\boldsymbol{E}_8, \boldsymbol{E}_7, \boldsymbol{E}_6$, or $\boldsymbol{F}_4$.

3. $p=5$ and $H=\boldsymbol{E}_8$.

An immediate consequence of Theorem 4 and this work of A. Borel is:

THEOREM 4′. *Let G be a compact Lie group of rank n, K a closed subgroup of* rank k, *and Γ a finite subgroup of the identity component G_0 of G such that $\Gamma \cap \operatorname{ad}(G)K = 1$. Then if p is a prime for which $\boldsymbol{H}^*(G_0, \boldsymbol{Z})$ has no p-torsion, every abelian subgroup of Γ has* p-rank $\leqq n-k$. *If $\boldsymbol{H}^*(G_0, \boldsymbol{Z})$ is torsion-free, every abelian subgroup of Γ has* rank $\leqq n-k$.

It is now clear that the problems in applying Theorem 4 are closely related to the existence of p-torsion in G. This is of two sorts—p-torsion from the fundamental group of G and p-torsion from the simply connected versions of the simple factors of G. Finally, we can only hope to classify our groups Γ in case rank. G — rank. $K \leqq 1$, due to the present rate of the theory of finite groups. We will see, however, that p-torsion in G is of little importance in case rank. $G =$ rank. K, and that only the p-torsion from $\pi_1(G)$ is of importance in case rank. G — rank. $K = 1$.

In addition to the results mentioned above, A. Borel has shown [2]

Let G be a compact connected Lie group, $\pi_1(G)$ torsion-free and $x \in G$. The centralizer of x in G is connected.

As the identity component of the centralizer of x in G is the union of the maximal tori of G which contain x, it easily follows, if $\pi_1(G)$ is torsion-free, that every abelian subgroup of G with 2 generators lies in a torus of G. We will depend heavily on this result of A. Borel in the next section.

II. 3. A further bound on the ranks of abelian subgroups. The main purpose of this section is to prove:

THEOREM 5. *Let G be a compact connected Lie group, K a closed subgroup with* rank. G — rank. $K = 1$, *Γ a finite subgroup of G with $\Gamma \cap \operatorname{ad}(G)K = 1$, p a prime, and $h(p)$ the* p-rank *of $\pi_1(G)$. Then every abelian subgroup of Γ has* p-rank $\leqq h(p)+1$. *If $h(p)=2$, then every abelian subgroup of Γ has* p-rank $\leqq 2$.

We will first need a few lemmas. The first two of these lemmas are known, but not well-known, so it seems best to write them out.

LEMMA 5.1. *Let G be a compact connected Lie group, G_{ss} the semi-*

simple part of G, and $Z(G)_0$ the identity component of the center of G. There is a covering $\phi\colon G_{ss}\times Z(G)_0\to G$ given by $\phi(g,t)=gt$. ϕ is an epimorphism of compact connected Lie groups and the kernel, ker. ϕ, *of ϕ is the set of all (g,g^{-1}) with $g\in G_{ss}\cap Z(G)_0$.*

Proof. G_{ss} has finite center and $G=G_{ss}\cdot Z(G)_0$. Note that ker. ϕ is finite and lies in the center of $G_{ss}\times Z(G)_0$.

LEMMA 5.2. *Let G be a compact connected Lie group. As a topological space, G is homeomorphic with $G_{ss}\times Z(G)_0$. As $Z(G)_0$ is a torus, it follows that the torsion subgroup of $\pi_1(G)$ is isomorphic to $\pi_1(G_{ss})$, and in particular* p-rank. $\pi_1(G)=p$-rank. $\pi_1(G_{ss})$ *for every prime p.*

Proof. We proceed by induction on the dimension s of the torus $Z(G)_0$, and the lemma is trivial if $s=0$. If $s=1$, we consider the principal bundle $G\to G/G_{ss}=Z(G)_0/(G_{ss}\cap Z(G)_0)$ with connected group G_{ss} and base which is a 1-sphere. Since this is a trivial bundle ([12], p. 99), G is homeomorphic to $G_{ss}\times$(1-sphere), which is homeomorphic to $G_{ss}\times Z(G)_0$. Now assume $s>1$. Take a subgroup H of G which is generated by G_{ss} and an $(s-1)$-torus T in $Z(G)_0$. H is homeomorphic to $G_{ss}\times T$ by induction. As before, the principal fibre bundle $G\to G/H$ tells us that G is homeomorphic to $H\times$(1-sphere), so G is homeomorphic to $G_{ss}\times Z(G)_0$.

Now note that $\pi_1(G)=\pi_1(G_{ss})\times\pi_1(\text{torus})$ and $\pi_1(\text{torus})$ is a free abelian group. *QED.*

LEMMA 5.3. *Let G be a compact connected Lie group, p a prime, and $h(p)$ the* p-rank *of $\pi_1(G)$. Let $\phi\colon G_{ss}\times Z(G)_0\to G$ be the covering given by $\phi(g,t)=gt$, $\mu\colon G'\to G_{ss}$ the universal covering of G_{ss}, and $\theta\colon G'\times Z(G)_0\to G$ the composition $\phi\cdot(\mu\times 1)$. Then every $(Z_p)^{h(p)+2}$ in G contains a $(Z_p)^2$ which is the θ-image of an abelian subgroup of $G'\times Z(G)_0$. If $h(p)=2$, every $(Z_p)^3$ in G contains a $(Z_p)^2$ which is the θ-image of an abelian subgroup of $G'\times Z(G)_0$.*

Proof. Let $\phi\beta_1,\cdots,\phi\beta_{h(p)+2}$ generate a $(Z_p)^{h(p)+2}$ in G, $N=$ ker. ϕ, and $\beta_j=(b_j,t_j)\in G_{ss}\times Z(G)_0$. As $[\beta_i,\beta_j]\in N$, where $[\ ,\ ]$ is the ordinary commutator, we know from the form of the elements of N that $[b_i,b_j]=1\in G_{ss}$. As $\beta_j{}^p\in N$ we also know that $b_j{}^p$ is central in G_{ss}. Now take elements $c_j\in G'$ with $\mu(c_j)=b_j$. As $b_j{}^p$ is central in G_{ss}, $c_j{}^p$ is central in G'. Since the b_j commute with each other, the commutators $[c_i,c_j]$ lies in ker. μ and are thus central in G'.

Let u and v be elements of a group H such that $w=[u,v]$ commutes

with u. $uv = wvu$ and we assume $u^{n-1}v = w^{n-1}vu^{n-1}$ by induction on n. Hence $u^n v = uu^{n-1}v = uw^{n-1}vu^{n-1} = w^{n-1}uvu^{n-1} = w^{n-1}wvuu^{n-1} = wv^n u^n$ in general. In other words, $[u^n, v] = [u, v]^n$ if u commutes with $[u, v]$. Since $[c_i, c_j]$ is central in G', it commutes with c_i, and consequently $[c_i, c_j]^p = [c_i^p, c_j]$ which equals 1 because c_i^p is central in G'.

$M = \ker.\mu$ is isomorphic to $\pi_1(G_{ss})$ so, by Lemma 5.2 and the definition of $h(p)$, M does not contain a $(Z_p)^{h(p)+1}$. Now set $y_j = [c_{h(p)+2}, c_j]$ for $1 \leqq j \leqq h(p)+1$. We have just seen that $y_j^p = 1$. As $\mu[c_i, c_j] = [\mu c_i, \mu c_j] = [b_i, b_j] = 1$, $y_j \in M$. It follows that the y_j generate an elementary p-subgroup Y in M of rank at most $h(p)$. Since there are $h(p)+1$ of the y_j, we have a relation $y_1^{v_1} y_2^{v_2} \cdots y_{h(p)+1}^{v_{h(p)+1}} = 1$, v_j integers not all divisible by p. Set $c = c_1^{v_1} c_2^{v_2} \cdots c_{h(p)+1}^{v_{h(p)+1}}$ and $t = t_1^{v_1} t_2^{v_2} \cdots t_{h(p)+1}^{v_{h(p)+1}}$ and notice that the fact that $[c_i, c_j c_k] = [c_i, c_j] \cdot [c_i, c_k]$, a consequence of the fact that each $[c_i, c_q]$ is central in G', gives us $[c_{h(p)+2}, c] = 1$. We now have elements $\sigma = (c, t)$ and $\tau = (c_{h(p)+2}, t_{h(p)+2})$ in $G' \times Z(G)_0$ such that σ and τ generate an abelian group in $G' \times Z(G)_0$ whose θ-image is a $(Z_p)^2$ inside our original $(Z_p)^{h(p)+2}$ in G.

Now suppose $h(p) = 2$. As before, we have a $(Z_p)^3$ in G generated by $\phi\beta_1, \phi\beta_2, \phi\beta_3$; we have $\beta_j = (b_j, t_j)$; and we have $\mu(c_j) = b_j$. We set $y_1 = [c_1, c_2]$, $y_2 = [c_2, c_3]$, $y_3 = [c_3, c_1]$ and the y_j generate an elementary p-subgroup of M which, by definition of $h(p)$, has rank $\leqq 2$. This gives us a relation of the form $y_1^{v_1} y_2^{v_2} y_3^{v_3} = 1$, where the v_j are integers not all divisible by p. We can assume that v_1 is not divisible by p, so there are integers r and s such that $y_1 = y_2^r y_3^s$. If p divides s, $[c_1 c_3^r, c_2] = 1$. If p doesn't divide s, there is an integer u with $us \equiv -r \pmod{p}$, and $[c_1 c_2^u, c_2 c_3^s] = 1$. In either case we get an abelian group from the c_j whose θ-image is a $(Z_p)^2$ inside our original $(Z_p)^3$ in G. *QED.*

Proof of Theorem 5. Let B be an abelian subgroup of Γ with p-rank.$B > h(p)+1$. Then B contains a $(Z_p)^{h(p)+2}$. By Lemma 5.3 we have a $(Z_p)^2$ in B which is the θ-image of an abelian subgroup S of $G' \times Z(G)_0$. S is generated by two elements. By a theorem of A. Borel, mentioned in § II.2, S lies in a torus of $G' \times Z(G)_0$, so $\theta(S)$ lies in a torus of G. Hence Γ contains a $(Z_p)^2$ which lies in a torus of G. As rank.G — rank.$K = 1$ and $\Gamma \cap \mathrm{ad}(G)K = 1$, this contradicts Theorem 4. The proof that $h(p) = 2$ implies p-rank.$B \leqq 2$ is identical. *QED.*

COROLLARY 5.1. *Let G be a compact connected Lie group which has torsion-free fundamental group, i.e., such that G_{ss} is simply connected. Let K be a closed subgroup of G such that* rank.G — rank.$K = 1$ *and let Γ be a*

finite subgroup of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$. *Then every abelian subgroup of* Γ *is cyclic. The odd Sylow subgroups of* Γ *are cyclic and the 2-Sylow subgroups are cyclic or generalized quaternionic, i. e., given by two generators* A *and* B *with the relations*

$$A^{2^{a-1}} = 1, \quad A^{2^{a-2}} = B^2, \quad BAB^{-1} = A^{-1}, \quad a \text{ integer}, \quad a > 2.$$

Proof. Let V be an abelian subgroup of Γ and write V as a product of p-subgroups. By Theorem 5, each of these p-subgroups has rank $\leqq 1$, hence is cyclic. Since V is a product of cyclic subgroups of pairwise relatively prime orders, it follows that V is cyclic. The rest is known ([14], Chapter I) to follow. *QED.*

Chapter III. An arithmetic criterion and first application to the classical groups.

III. 1. Angular parameters and the arithmetic criterion. Let G be a compact connected Lie group of rank n, K a closed connected subgroup of rank k, T a maximal torus of G which contains a maximal torus T' of K, $W = \{w_1, \cdots, w_q\}$ an enumeration of the Weyl group of G relative to T, and $T_i = w_i(T')$. We choose an *integral basis* of the Lie algebra T of T, i. e., an ordered basis $\mathsf{X} = \{X_1, \cdots, X_n\}$ of T such that $\exp(\sum_s a_s X_s) = 1$ if and only if each a_s is an integer. The Lie algebra T_i of T_i can be described as the set of all elements $\sum_s a_s X_s$ of T such that $\sum_s v_{ijs} a_s = 0$ for $1 \leqq j \leqq n - k$, where each $\{v_{ij1}, \cdots, v_{ijn}\} = V_{ij}$ is an ordered set of relatively prime integers. The v_{ijs} can be chosen to be rational because each T_i is closed in T, and the obvious normalization transforms each V_{ij} into a set of relatively prime integers.

DEFINITION. *The* $q(n-k)$ *ordered sets* V_{ij} *of relatively prime integers are the angular parameters of* K *in* G *relative to* X.

We remark that, for a given K and G, the choice of X does not specify the angular parameters of K in G uniquely.

Let Γ be a finite subgroup of G. We choose cyclic subgroups $\{\gamma_t\} = \Gamma_t$ of Γ such that every element of Γ is $\mathrm{ad}(G)$-conjugate to an element of one of the Γ_t. Then $\Gamma \cap \mathrm{ad}(G)K = 1$ if and only if $\Gamma_t \cap \mathrm{ad}(G)K = 1$ for each t. The angular parameters of K in G give us an arithmetic criterion for $\Gamma_t \cap \mathrm{ad}(G)K = 1$:

THEOREM 6. *Let G be a compact connected Lie group of* rank n, *K a*

closed connected subgroup of rank k, $V_{ij}=\{v_{ij1},v_{ij2},\cdots,v_{ijn}\}$ *the angular parameters of* K *in* G *relative to an integral basis* $\mathsf{X}=\{X_1,\cdots,X_n\}$ *of the Lie algebra* T *of a maximal torus of* G, $B=\{\beta\}$ *a cyclic subgroup of order* m *in* G, *and* $X=\sum_s a_sX_s\in\mathsf{T}$ *such that* $\exp(X)$ *is* $\mathrm{ad}(G)$*-conjugate to* β. *Then each* $b_s=ma_s$ *is an integer, and* $B\cap\mathrm{ad}(G)K=1$ *if and only if each*

$$V_i=\{m,\textstyle\sum_s v_{i1s}b_s,\sum_s v_{i2s}b_s,\cdots,\sum_s v_{i\,n-k\,s}b_s\}$$

is a set of relatively prime integers.

Proof. Each $b_s=ma_s$ is an integer because X is an integral basis of T and $\exp(mX)=\exp(\sum_s b_sX_s)$ is conjugate to $\beta^m=1$.

We will use the notation leading to the definition of the angular parameters of K in G relative to X. An element of T lies in $\mathrm{ad}(G)K$ if and only if it lies in one of the T_i, so $B\cap\mathrm{ad}(G)K=1$ if and only if $\exp(rX)\notin T_i$ for any i whenever $r\not\equiv 0\ (\mathrm{mod}\,m)$. X being an integral basis of T, $\exp(rX)\in T_i$ if and only if there is a choice a_{is} of integers such that $rX+\sum_s a_{is}X_s=\sum_s(ra_s+a_{is})X_s$ lies in T_i, i.e., such that $\sum_s v_{ijs}(rb_s+ma_{is})=0$ for $1\leqq j\leqq n-k$. Reducing modulo m this says that $r\sum_s v_{ijs}b_s\equiv 0$ for $1\leqq j\leqq n-k$. If $r\not\equiv 0\ (\mathrm{mod}\,m)$, this implies that V_i is not a set of relatively prime integers.

Now suppose that V_i is not a set of relatively prime integers. Then there is an integer $r\not\equiv 0\ (\mathrm{mod}\,m)$ such that $r\sum_s v_{ijs}b_s\equiv 0\ (\mathrm{mod}\,m)$ for $1\leqq j\leqq n-k$. We will show that $\exp(rX)\in T_i$, so $\beta^r\in\mathrm{ad}(G)K$. Let U_{ij} be the $(n-1)$-torus whose Lie algebra U_{ij} is the hyperplane $\sum_s v_{ijs}x_s=0$ in T, where the x_s are coordinates relative to the basis X. $T_i=\bigcap_j U_{ij}$. V_{ij} being a set of relatively prime integers, we have integers c_{ijs} with $\sum_s c_{ijs}v_{ijs}=1$. The congruences $r\sum_s v_{ijs}b_s\equiv 0\ (\mathrm{mod}\,m)$ gives us integers t_{ij} with $mt_{ij}+r\sum_s v_{ijs}b_s=0$, so we have integers $a_{ijs}=c_{ijs}t_{ij}$ such that $\sum_s v_{ijs}(rb_s+ma_{ijs})=0$, for $1\leqq j\leqq n-k$. This just says that $\exp(rX)\in U_{ij}$ for $1\leqq j\leqq n-k$, so $\exp(rX)\in T_i$. *QED.*

Theorem 6 can be used to check $\Gamma\cap\mathrm{ad}(G)K=1$ provided that the intersection of K with the identity component G_0 of G is connected and $\Gamma\subset G_0$. Let $\{f_t\}$ be automorphisms of G such that the automorphism group $\mathrm{ad}(G)$ of G is the union of the $\mathrm{ad}(G_0)\cdot f_t$. Let $K_t=f_t(K)\cap G_0$. Given $\Gamma\subset G_0$, $\Gamma\cap\mathrm{ad}(G)K=1$ if and only if $\Gamma\cap\mathrm{ad}(G_0)K_t=1$ for every t. We can check the $\Gamma\cap\mathrm{ad}(G_0)K_t=1$ with Theorem 6 and thus check $\Gamma\cap\mathrm{ad}(G)K=1$.

The application of Theorem 6 is simplified when the Weyl group W of G acts on the integral basis X by signed permutations: the angular parameters can then be chosen so that each V_{ij} is obtained from V_{1j} by the same signed permutations. We will use this trick when G is a classical group.

III. 2. Even and odd subgroups of the classical groups. By the *classical groups* we mean the unitary groups $\boldsymbol{U}(n)$, the special unitary groups $\boldsymbol{SU}(n)$, the symplectic groups (often called the unitary symplectic groups) $\boldsymbol{Sp}(n)$, the special orthogonal groups $\boldsymbol{SO}(n)$, and the spin groups (universal covering groups of the special orthogonal groups) $\boldsymbol{Spin}(n)$. They are all compact connected Lie groups. $\boldsymbol{U}(n)$ has rank n, semisimple part $\boldsymbol{SU}(n)$ and fundamental group infinite cyclic. $\boldsymbol{SU}(n+1)$ has rank n, is simple for $n \geqq 1$, and is simply connected. $\boldsymbol{Sp}(n)$ has rank n, is simple for $n \geqq 1$, and is simply connected. $\boldsymbol{Sp}(n)$ can be viewed as all elements of $\boldsymbol{U}(2n)$ which preserve an antisymmetric nondegenerate 2-form on complex Euclidean space C^{2n}. Given an orthonormal basis $\{e_1, \cdots, e_{2n}\}$ of C^{2n}, we'll use the form

$$A\left(\sum_1^{2n} x_j e_j, \sum_1^{2n} y_j e_j\right) = \sum_1^{n} (x_j y_{j+n} - y_j x_{j+n}).$$

$\boldsymbol{SO}(2n \text{ or } 2n+1)$ has rank n. $\boldsymbol{SO}(k)$ is semisimple for $k \geqq 3$, simple for $4 \neq k \geqq 3$, has fundamental group Z_2 if $k \geqq 3$, and has universal covering group $\boldsymbol{Spin}(k)$.

Given a classical group G, we have a canonical choice of a maximal torus T of G:

1. $G = \boldsymbol{U}(n)$. T is the set of all matrices $\operatorname{diag}\{a_1, \cdots, a_n\}$, where each a_j is a unimodular complex number.

2. $G = \boldsymbol{SU}(n)$. T is the set of all $\operatorname{diag}\{a_1, \cdots, a_n\}$ of determinant 1, where each a_j is a unimodular complex number.

3. $G = \boldsymbol{Sp}(n)$. T is the set of all matrices $\begin{pmatrix} D & 0 \\ 0 & \bar{D} \end{pmatrix}$, where $\bar{D}$ is the complex conjugate of D and D is in the canonical maximal torus of $\boldsymbol{U}(n)$.

4. $G = \boldsymbol{SO}(2n \text{ or } 2n+1)$. T is the set of all matrices $\operatorname{diag}\{R(t_1), \cdots, R(t_n), (1)\}$, where $R(t) = \begin{pmatrix} \cos 2\pi t & \sin 2\pi t \\ -\sin 2\pi t & \cos 2\pi t \end{pmatrix}$ and the (1) appears only if $G = \boldsymbol{SO}(2n+1)$.

5. $G = \boldsymbol{Spin}(2n \text{ or } 2n+1)$. T is the complete inverse image of our chosen maximal torus in $\boldsymbol{SO}(2n \text{ or } 2n+1)$.

If G is not a ***Spin*** or special unitary group, we have a canonical choice of integral basis $\mathsf{X}_G = \{X_1, \cdots, X_n\}$ of the Lie algebra T of T:

1. $G = \boldsymbol{U}(n)$. $\exp(tX_j) = \operatorname{diag}\{1, \cdots, 1, \exp(2\pi i t), 1, \cdots, 1\}$, where the $\exp(2\pi i t)$ is in the j-th place and $i^2 = -1$.

2. $G = \boldsymbol{Sp}(n)$. $\exp(tX_j) = \begin{pmatrix} D & 0 \\ 0 & \bar{D} \end{pmatrix}$, where $\bar{D}$ is the complex conjugate of D and $D = \operatorname{diag}\{1, \cdots, 1, \exp(2\pi it), 1, \cdots, 1\}$ has the $\exp(2\pi it)$ in the j-th place.
3. $G = \boldsymbol{SO}(2n$ or $2n+1)$. Let I_2 be the 2×2 identity matrix; $\exp(tX_j) = \operatorname{diag}\{I_2, \cdots, I_2, R(t), I_2, \cdots, I_2, (1)\}$, where $R(t)$ is the j-th block.

The Weyl group W of G acts on X_G by signed permutations:

1. $G = \boldsymbol{U}(n)$. W acts on X_G by all permutations.
2. $G = \boldsymbol{Sp}(n)$ or $\boldsymbol{SO}(2n+1)$. W acts on X_G by all signed permutations.
3. $G = \boldsymbol{SO}(2n)$. W acts on X_G by all signed permutations where the the number of sign changes is even.

Let K be a closed connected subgroup of rank k in a classical group $G = \boldsymbol{U}(n)$, $\boldsymbol{Sp}(n)$ or $\boldsymbol{SO}(2n$ or $2n+1)$. Replacing K by a conjugate, we have a maximal torus T' of K which lies in our canonical maximal torus T of G. The Lie algebra T' of T' is the intersection of $n-k$ hyperplanes $\sum_s v_{js} x_s = 0$, where the x_s are coordinates in T relative to the canonical integral basis X_G; we can assume that each $V_j = \{v_{j1}, v_{j2}, \cdots, v_{jn}\}$ is a set of relatively prime integers. If $W = \{w_1, \cdots, w_q\}$ is an enumeration of the Weyl group of G relative to T, W envisaged as a group of signed permutations on n-tuples from its action on X_G, the angular parameters of K in G relative to X_G are the $V_{ij} = w_i(V_j)$.

DEFINITION. *The $n-k$ ordered sets V_j of relatively prime integers are the canonical parameters of K in $G = \boldsymbol{U}(n)$, $\boldsymbol{Sp}(n)$ or $\boldsymbol{SO}(2n$ or $2n+1)$.*

Let K be a closed connected subgroup of rank $k-1$ in $\boldsymbol{SU}(n)$. Viewing $\boldsymbol{SU}(n)$ as a subgroup of $\boldsymbol{U}(n)$, K has canonical parameters $V_1, \cdots, V_{n-k+1}$ in $\boldsymbol{U}(n)$. We may assume that $V_{n-k+1} = \{1, 1, \cdots, 1\}$.

DEFINITION. *The $n-k$ ordered sets $V_1, \cdots, V_{n-k}$ of relatively prime integers are the canonical parameters of K in $\boldsymbol{SU}(n)$.*

Let K be a closed connected subgroup of rank k in $\boldsymbol{Spin}(2n$ or $2n+1)$ and let $f\colon \boldsymbol{Spin} \to \boldsymbol{SO}$ be the natural projection. We will use the canonical parameters $V_1, \cdots, V_{n-k}$ of $f(K)$ in $\boldsymbol{SO}(2n$ or $2n+1)$ for the canonical parameters of K in $\boldsymbol{Spin}(2n$ or $2n+1)$:

DEFINITION. *The $n-k$ ordered sets V_j of relatively prime integers are the canonical parameters of K in $\boldsymbol{Spin}(2n$ or $2n+1)$.*

Given a closed connected subgroup K of a classical group G and an integral basis X of the Lie algebra of a maximal torus of G, we can always construct the angular parameters of K in G relative to X from the canonical parameters of K in G.

The fact that the Weyl group acts on the canonical parameters by signed permutations allows us to define:

Definition. *Let K be a closed connected subgroup of a classical group $G = \boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(2n$ or $2n+1)$ or $\boldsymbol{Spin}(2n$ or $2n+1)$ such that* rank. G — rank. $K = 1$. *Let $V = \{v_1, \cdots, v_n\}$ be the canonical parameter of K in G and set $v = v_1 v_2 \cdots v_n$. Then K is an even subgroup of G if v is an even integer; K is an odd subgroup of G if v is an odd integer.*

The most familiar examples of even subgroups are

$$\boldsymbol{U}(n-1) \subset \boldsymbol{U}(n),\ \boldsymbol{SU}(n-1) \subset \boldsymbol{SU}(n),\ \boldsymbol{Sp}(n-1) \subset \boldsymbol{Sp}(n),$$
$$\boldsymbol{SO}(2n-1) \subset \boldsymbol{SO}(2n) \text{ and } \boldsymbol{Spin}(2n-1) \subset \boldsymbol{Spin}(2n).$$

In these examples the canonical parameter can be taken to be $\{1, 0, \cdots, 0\}$.

III. 3. The orthogonal groups. If G is a classical group $\boldsymbol{U}(n)$, $\boldsymbol{SU}(n+1)$, $\boldsymbol{Sp}(n)$ or $\boldsymbol{Spin}(2n$ or $2n+1)$ of rank n, K is a closed connected subgroup of rank $n-1$ and Γ is a finite subgroup of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$, then Corollary 5.1 tells us that every abelian subgroup of Γ is cyclic. If, however, $G = \boldsymbol{SO}(2n$ or $2n+1)$, then we only know that every abelian subgroup of Γ is of the form $Z_u \times Z_v$, where u is a power of 2. As it is known [11] that a $(Z_2)^2$ cannot act freely on the sphere $S^{2n-1} = \boldsymbol{SO}(2n)/\boldsymbol{SO}(2n-1)$, there is, at least for some choices of K, room for improvement:

Theorem 7. *Let G be a special orthogonal group $\boldsymbol{SO}(q) = \boldsymbol{SO}(2n$ or $2n+1)$ of rank n and let K be a closed connected subgroup of rank $n-1>0$. If K is odd, G has a subgroup B isomorphic to $(Z_2)^2$ with $B \cap \mathrm{ad}(G)K = 1$. If K is even and Γ is a finite subgroup of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$, then every abelian subgroup of Γ is cyclic.*

Proof. Let $V = \{v_1, \cdots, v_n\}$ be the canonical parameter of K in G and let $b \in G$ have order 2. The eigenvalues of b are all 1 or -1. As det. $b = 1$, the multiplicity of the eigenvalue -1 is some even number $2s$. It is clear that b is $\mathrm{ad}(G)$-conjugate to $\exp(\frac{1}{2}X_1 + \frac{1}{2}X_2 + \cdots + \frac{1}{2}X_s)$, where $\mathsf{X}_G = \{X_1, \cdots, X_n\}$ is our canonical integral basis, so the arithmetic criterion (Theorem 6) says that $b \in \mathrm{ad}(G)K$ if and only if some sum of s of the v_j, without repetitions, is even. When the v_j are all odd this means that

$b \in \mathrm{ad}(G)K$ if and only if s is even; when one of the v_j is even and $s < n$, this implies that $b \in \mathrm{ad}(G)K$.

Suppose K is odd. Then each of the v_j is odd, so we must exhibit a $(Z_2)^2$ in G in which every element $\neq 1$ has the eigenvalue -1 of multiplicity congruent to 2 modulo 4. Let $\boldsymbol{I}_t$ be the $t \times t$ identity matrix; then such a $(Z_2)^2$ is given by generators

$$b_1 = \mathrm{diag}\{-1, -1, 1, \boldsymbol{I}_{q-3}\}, \qquad b_2 = \mathrm{diag}\{1, -1, -1, \boldsymbol{I}_{q-3}\}$$

Suppose that K is even, so one of the v_j is even. By Theorem 5 we need only show that Γ contains no $(Z_2)^2$, so we must show that a $(Z_2)^2$ in G has an element $\neq 1$ with eigenvalue -1 of multiplicity not equal to $2n$. A $(Z_2)^m$ in $\boldsymbol{SO}(q)$ is conjugate to a group of diagonal matrices. It follows that G contains a $(Z_2)^2$ where every element $\neq 1$ has eigenvalue -1 with multiplicity $2n$ only if $q = 3$. That case was ruled out by the assumption rank. $K > 0$. *QED.*

III. 4. Elements of order 2 which act freely.

THEOREM 8. *Let G be a classical group $\boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(2n)$ or $\boldsymbol{Spin}(2n)$ and let K be an even subgroup (hence closed and connected, and* rank. G — rank. $K = 1$). *Let Γ be a finite subgroup of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$. Then Γ has at most one element of order 2, and an element of order 2 in Γ is central in G. Let H be a closed connected subgroup of $\boldsymbol{SU}(n)$ such that* rank. $\boldsymbol{SU}(n)$ — rank. $H = 1$ *and let Σ be a finite subgroup of $\boldsymbol{SU}(n)$ such that $\Sigma \cap \mathrm{ad}(\boldsymbol{SU}(n))H = 1$. Then both n and H are even if Σ has an element of order 2.*

Proof. Suppose $G \neq \boldsymbol{Spin}(2n)$ and let $\gamma \in \Gamma$ have order 2. As in the proof of Theorem 7, the arithmetic criterion shows that γ has the eigenvalue -1 with multiplicity $2n$ if $G = \boldsymbol{SO}(2n)$ or $\boldsymbol{Sp}(n)$, and with multiplicity n if $G = \boldsymbol{U}(n)$ or $\boldsymbol{SU}(n)$. Hence γ is conjugate to $-\boldsymbol{I}$, the negative of the identity matrix in G. As $-\boldsymbol{I}$ is central in G, $\gamma = -\boldsymbol{I}$ and is central in G.

Now suppose that $G = \boldsymbol{Spin}(2n)$ and $f \colon G \to \boldsymbol{SO}(2n)$ is the natural map. Let -1 denote the element of order 2 in ker. f. If -1 is in Γ or K, then Γ or K consists of whole f-fibres and we have $f(\Gamma) \cap \mathrm{ad}(\boldsymbol{SO}(2n))f(K) = \boldsymbol{I}_{2n}$. If -1 is in neither Γ nor K, then either $f(\Gamma) \cap \mathrm{ad}(\boldsymbol{SO}(2n))f(K) = \boldsymbol{I}_{2n}$ or Γ has an element $\gamma \neq 1$ such that $\boldsymbol{I}_{2n} \neq f(\gamma) \in f(\Gamma) \cap \mathrm{ad}(\boldsymbol{SO}(2n))f(K)$. We will show that this last alternative does not occur. For if it does, K has a conjugate K' such that $-\gamma \in K'$. γ has order 2, for $\gamma \notin$ ker. f but $\gamma^2 = (-\gamma)^2 \in \Gamma \cap K'$. We can pass to a conjugate of γ and assume

$\gamma = e_1 \cdot e_2 \cdot \cdots \cdot e_{2s}$, where the e_j are an orthonormal basis of Euclidean space $\boldsymbol{R}^{2n}$, taken as generators of the Clifford algebra $C(\boldsymbol{R}^{2n})$, and dots denote Clifford multiplication. If $s = n$, γ is central and thus, by Theorem 5, the only element of Γ of order 2. If $s < n$, let $\beta = e_{2s} \cdot e_{2s+1} \in \boldsymbol{Spin}(2n)$ and $\mathrm{ad}(\beta)\gamma = -\gamma$. This implies that both γ and $-\gamma$ are in $\mathrm{ad}(G)K$, which contradicts $\Gamma \cap \mathrm{ad}(G)K = 1$. Now we can assume that $f(\Gamma) \cap \mathrm{ad}(\boldsymbol{SO}(2n))f(K) = \boldsymbol{I}_{2n}$. $f(K)$ is an even subgroup of $\boldsymbol{SO}(2n)$ because K is even in $\boldsymbol{Spin}(2n)$, so an element of $f(\Gamma)$ of order 2 is $-\boldsymbol{I}_{2n}$. It follows that an element of Γ of order 2 lies in $f^{-1}(\{\pm \boldsymbol{I}_{2n}\})$, hence is central in $\boldsymbol{Spin}(2n)$. Uniqueness follows from Theorem 5.

Let $G = \boldsymbol{SU}(n)$ and let $\sigma \in \Sigma$ have order 2. If H is odd, the arithmetic criterion implies that the eigenvalue -1 of σ has odd multiplicity, contradicting $\det. \sigma = 1$. Thus H is even. If n is odd, we again contradict $\det. \sigma = 1$, for, H being even, the arithmetic criterion says that $\sigma = -\boldsymbol{I}_n$. *QED.*

Chapter IV. Finite subgroups of classical groups which have all abelian subgroups cyclic.

Theorems 5 and 7 tell us that if G is a classical group and K is a closed connected subgroup, assumed to be an even subgroup if G is special orthogonal, such that $\mathrm{rank}. G - \mathrm{rank}. K = 1$, and Γ is a finite subgroup of G such that $\Gamma \cap \mathrm{ad}(G)K = 1$, then every abelian subgroup of Γ is cyclic. For this reason, we'll examine the finite groups with all abelian subgroups cyclic.

IV. 1. Classification of finite groups with all abelian subgroups cyclic. The finite groups with all abelian subgroups cyclic fall into two classes ([14], Chapter I)—those with all Sylow subgroups cyclic, and those with odd Sylow subgroups cyclic and 2-Sylow subgroups generalized quaternionic. H. Zassenhaus ([16], p. 198, p. 202) and M. Suzuki ([13], p. 689) have given a complete classification of these groups in terms of generators and relations. We will not use this classification, but rather will rely on a simpler description given in H. Zassenhaus' book ([17], p. 175) for the finite groups with all Sylow subgroups cyclic, and on the fact that a finite group with all abelian subgroups cyclic has all Sylow subgroups cyclic if its 2-Sylow subgroups are not generalized quaternionic. For reference, the generalized quaternionic groups are the groups $\boldsymbol{Q}2^a$ of order 2^a, $a \geqq 3$, given by

$$A^{2^{a-1}} = 1, \quad B^2 = A^{2^{a-2}}, \quad BAB^{-1} = A^{-1}, \quad a \text{ integer}, \quad a \geqq 3.$$

A finite group of order N with all Sylow subgroups cyclic is given by $A^m = B^n = 1$, $BAB^{-1} = A^r$, $0 < m$, $mn = N$, $((r-1)n, m) = 1$, $r^n \equiv 1 \pmod{m}$.

Our plan of attack is to calculate representations of these groups in the classical groups, and find conditions under which the image of a representation acts freely on an appropriate coset space.

IV. 2. Classical representations of the generalized quaternionic groups. Following G. Vincent ([14], Chapter III), elementary techniques of representation theory tell us that the irreducible unitary representations of the generalized quaternionic group $\boldsymbol{Q}2^a$ are:

1. The 4 $\boldsymbol{U}(1)$-representations given by $A \to \pm 1$ and $B \to \pm 1$.
2. The $2^{a-2}-1$ $\boldsymbol{U}(2)$-representations S_r, $1 \leqq r < 2^{a-2}$, given by

$$S_r\colon A \to \begin{pmatrix} u^r & 0 \\ 0 & u^{-r} \end{pmatrix} \text{ and } B \to \begin{pmatrix} 0 & 1 \\ (-1)^r & 0 \end{pmatrix}, \text{ where } u = \exp(2\pi i/2^{a-1}).$$

Note that S_r is faithful if and only if r is odd. Let S denote the $\boldsymbol{U}(1)$- representation $A \to 1$ and $B \to -1$.

It follows that a special unitary representation of $\boldsymbol{Q}2^a$ is an appropriate sum of $\boldsymbol{U}(1)$-representations plus a sum of some of

1. The 2^{a-3} $\boldsymbol{SU}(2)$-representations S_r, where r is odd.
2. The 2^{a-3} $\boldsymbol{SU}(3)$-representations $S_r + S$, where r is even.
3. The $\boldsymbol{SU}(4)$-representations $S_{r_1} + S_{r_2}$, where r_1 and r_2 are even.

Similarly, a symplectic representation of $\boldsymbol{Q}2^a$ is an appropriate sum of $\boldsymbol{U}(1)$-representations plus a sum of some of

1. The 2^{a-3} $\boldsymbol{Sp}(1)$-representations S_r, where r is odd.
2. The 2^{a-3} $\boldsymbol{Sp}(1)$-representations $S_r + S_r^*$, where r is even and S_r^* is the complex conjugate representation of S_r.

A unitary, special unitary or symplectic representations of $Q2^a$ is faithful if and only if it has a summand S_r with r odd.

S_r is unitarily equivalent to its conjugate representation S_r^*, and is equivalent to a real representation if and only if r is even. As before, we set $R(t) = \begin{pmatrix} \cos(2\pi t) & \sin(2\pi t) \\ -\sin(2\pi t) & \cos(2\pi t) \end{pmatrix} \in \boldsymbol{SO}(2)$; the irreducible orthogonal representations of $\boldsymbol{Q}2^a$ are:

1. The 4 $\boldsymbol{O}(1)$-representations given by $A \to \pm 1$ and $B \to \pm 1$.

2. The $2^{a-3}-1$ $\boldsymbol{O}(2)$-representations S_r, r even, unitarily equivalent to the corresponding $\boldsymbol{U}(2)$-representations, given by S_r: $A\to R(r/2^{a-1})$ and $B\to\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$

3. The 2^{a-3} $\boldsymbol{O}(4)$-representations T_r, r odd, equivalent to $S_r+S_r{}^*$, given by T_r: $A\to\begin{pmatrix}R(r/2^{a-1}) & 0\\ 0 & R(-r/2^{a-1})\end{pmatrix}$ and $B\to\begin{pmatrix}0 & I_2\\ -I_2 & 0\end{pmatrix}$.

A special orthogonal representation of $\boldsymbol{Q}2^a$ is an appropriate sum of $\boldsymbol{O}(1)$-representation plus a sum of some of

1. The $\boldsymbol{SO}(3)$-representations S_r+S (r even, of course).
2. The $\boldsymbol{SO}(4)$-representations $S_{r_1}+S_{r_2}$ with r_1 and r_2 even.
3. The $\boldsymbol{SO}(4)$-representations T_r (r odd, of course).

An orthogonal or special orthogonal representation of $\boldsymbol{Q}2^a$ is faithful if and only if it has a summand T_r.

Each of the T_r can be lifted to a faithful $\boldsymbol{Spin}(4)$-representation of $\boldsymbol{Q}2^a$. Let T be one of the T_r and let $\{e_j\}$ be the orthonormal basis of $\boldsymbol{R}^4$ with respect to which our matrices are written; the $\{e_j\}$ generate the Clifford algebra $\boldsymbol{C}(\boldsymbol{R}^4)$. We choose $T'(A)\in\boldsymbol{Spin}(4)$ over $T(A)$ and $T'(B)\in\boldsymbol{Spin}(4)$ over $T(B)$. We then have

$$T'(A)=\pm(\cos x+e_2\cdot e_1\sin x)\cdot(\cos x-e_4\cdot e_3\sin x),$$

$$T'(A)^{-1}=\pm(\cos x-e_2\cdot e_1\sin x)\cdot(\cos x+e_4\cdot e_3\sin x)$$

and

$$T'(B)=\pm\tfrac{1}{2}(1+e_3\cdot e_1)\cdot(1+e_4\cdot e_2),$$

where dots denote Clifford multiplication, and $x=\pi r/2^{a-1}$. A short calculation shows that $T'(A)^{2^{a-2}}=e_1\cdot e_2\cdot e_4\cdot e_3=T'(B)^2$. Another calculation shows that $T'(B)\cdot T'(A)=T'(A)^{-1}\cdot T'(B)$. It follows that $T'(A)$ and $T'(B)$ generate a $\boldsymbol{Q}2^a$ in $\boldsymbol{Spin}(4)$, so T' extends to a $\boldsymbol{Spin}(4)$-representation of $\boldsymbol{Q}2^a$. T' is faithful because it covers a faithful representation.

Let $V=S_{r_1}+S_{r_2}$, $r_i=2u_i$, a non-faithful $\boldsymbol{SO}(4)$-representation of $\boldsymbol{Q}2^a$. If $V'(A)\in\boldsymbol{Spin}(4)$ lies over $V(A)$ and $V'(B)\in\boldsymbol{Spin}(4)$ lies over $V(B)$, a short calculation shows that $V'(B)^2=-1$, $V'(B)\cdot V'(A)\cdot V'(B)^{-1}=V'(A)^{-1}$, and $V'(A)^{2^{a-2}}=-1$ if and only if u_1+u_2 is odd. In other words, V' extends to a $\boldsymbol{Spin}(4)$-representation $V'$of $\boldsymbol{Q}2^a$ if and only if one of the u_i is odd and the other is even. In that case, V' is faithful and -1 is the element of order 2 in $V'(\boldsymbol{Q}2^a)$.

Let $U=S_r+S$, $r=2u$, a non-faithful $\boldsymbol{SO}(3)$-representation of $\boldsymbol{Q}2^a$.

Choosing $U'(A)$ and $U'(B)$ in $\boldsymbol{Spin}(3)$ over $U(A)$ and $U(B)$, we see that $U'(B)^2 = -1$, $U'(B)\cdot U'(A)\cdot U'(B)^{-1} = U'(A)^{-1}$, and $U'(A)^{2^{a-2}} = -1$ if and only if u is odd. Thus U' extends to a $\boldsymbol{Spin}(3)$-representation U' of $\boldsymbol{Q}2^a$ if and only if u is odd; in that case, U' is faithful and -1 is the element of order 2 in $U'(\boldsymbol{Q}2^a)$.

IV. 3. Unitary representations of finite groups which have all Sylow subgroups cyclic. Let Γ be a finite group of order N with every Sylow subgroup cyclic. Γ is given by two generators A and B with relations $A^m = B^n = 1$, $BAB^{-1} = A^r$, $0 < m$, $mn = N$, $((r-1)n, m) = 1$, $r^n \equiv 1 \pmod{m}$. Note that m is odd; if m were even, r would be odd because A and A^r have the same order, so $2 \mid ((r-1)n, m)$, where we denote *a divides b* by $a \mid b$. Note also that not r but only the mod m residue class of r is important. Let ϕ be the Euler ϕ-function and let G_m be the multiplicative group of integers prime to m, taken modulo m. As m is odd, there can be no confusion with the exceptional Lie group $\boldsymbol{G}_2$. Given $C \in \Gamma$, let $\{C\}$ denote the cyclic subgroup of Γ generated by C. Let d be the order of r in G_m. As $r^n \equiv 1 \pmod{m}$, $d \mid n$ and we can write $n = n'd$. If $m_i \mid m$, set $d_i =$ order of r in G_{m_i}, $n = d_i n_i'$. G. Vincent has proved ([14], p. 156):

Γ has exactly $\phi(m_j)n'd/d_j^2$ irreducible unitary representations of degree d_j. On restricting one of these representations to $\{A\}$, it has kernel $\{A^{m_j}\}$. As m_j runs over all divisors of m, including 1 and m, these representations of degree d_j give all irreducible unitary representations of Γ.

This, together with other results of G. Vincent ([14], Chapter III) make it fairly easy to verify that the irreducible unitary representations of degree d_j of Γ are given by:

$$f_{jk}(A) = \operatorname{diag}\{\exp(2\pi i/m_j), \exp(2\pi i r/m_j), \cdots, \exp(2\pi i r^{d_j-1}/m_j)\}$$

$$f_{jk}(B) = \exp(2\pi i k/n)\cdot\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \text{ for } 1 \leq k \leq \phi(m_j)n'd/d_j^2.$$

Chapter V. A second application of the arithmetic criterion to the classical groups.

V. 1. Generalized quaternionic subgroups of classical groups which act freely. Suppose G is a classical group of rank n and K is a closed

connected subgroup of rank $n-1$, assumed to be an even subgroup if G is a special orthogonal group. A generalized quaternionic subgroup Γ of G can be considered to be the image of a faithful G-representation of the appropriate $\boldsymbol{Q}2^a$. Since we know these representations explicitly, we can apply the arithmetic criterion to check whether $\Gamma \cap \mathrm{ad}(G)K = 1$. The importance of this procedure is that if no generalized quaternionic subgroup of G can act freely on G/K, then every finite subgroup of G that acts freely on G/K has all Sylow subgroups cyclic. Such groups have particularly simple structure among the finite groups with all abelian subgroups cyclic.

THEOREM 9.1. *Let G be a classical group $\boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, or $\boldsymbol{SO}(2n$ or $2n+1)$, K a closed connected subgroup with* rank. G — rank. $K = 1$, *$V = \{v_1, \cdots, v_n\}$ the canonical parameter of K in G, and q the number of v_i which are odd. Then G has a generalized quaternionic subgroup Γ such that $\Gamma \cap \mathrm{ad}(G)K = 1$ if and only if*

n is even and q is odd, if $G \neq \boldsymbol{Sp}(n)$; q is odd or $q = n$, if $G = \boldsymbol{Sp}(n)$. If G has a generalized quaternionic subgroup Γ such that $\Gamma \cap \mathrm{ad}(G)K = 1$ and if B is any generalized quaternionic group, then G has a subgroup B' isomorphic to B such that $B' \cap \mathrm{ad}(G)K = 1$. If K is an even subgroup in case $G = \boldsymbol{SO}(2n$ or $2n+1)$, if

n is odd or q is even, if $G \neq \boldsymbol{Sp}(n)$; q is even and $q < n$, if $G = \boldsymbol{Sp}(n)$, and if Σ is a finite subgroup of G such that $\Sigma \cap \mathrm{ad}(G)K = 1$, then every Sylow subgroup of Σ is cyclic.

Proof. Let Γ be a generalized quaternionic subgroup of $G = \boldsymbol{U}(n)$, $\boldsymbol{Sp}(n)$ or $\boldsymbol{SO}(2n$ or $2n+1)$, considered as the image of a faithful G-representation F of $\boldsymbol{Q}2^a$. Checking the various possible summands of F, we see that every element of Γ is $\mathrm{ad}(G)$-conjugate to a power of $F(A)$ or $F(B)$; it follows that $\Gamma \cap \mathrm{ad}(G)K = 1$ if and only if $\{F(A)\} \cap \mathrm{ad}(G)K = 1 = \{F(B)\} \cap \mathrm{ad}(G)K$. We will apply the arithmetic criterion (Theorem 6) to these two cyclic subgroups of Γ. When we do this, the integer m in the formulation of the arithmetic criterion will be a power of 2, as Γ is a 2-group, so we may ignore the integers b_i which are even. As the kernel of a non-faithful G-representation of $\boldsymbol{Q}2^a$ contains the element $B^2 = A^{2^{a-2}}$ of order 2, this means we need only consider the faithful summands of F. Replacing Γ by an $\mathrm{ad}(G)$-conjugate if necessary, we can assume $F = S_{r_1} + \cdots + S_{r_s}$ if $G = \boldsymbol{U}(n)$ or $\boldsymbol{Sp}(n)$ and $F = T_{r_1} + \cdots + T_{r_s}$ if $G = \boldsymbol{SO}(2n$ or $2n+1)$, where the r_i are odd. An application of the arithmetic criterion now tells us that the condtion for $\Gamma \cap \mathrm{ad}(G)\mathrm{K} = 1$ is that, for every element g of the

group $S(n)$ of all permutations on $\{1,2,\cdots,n\}$, $\sum_j r_j(v_{g(2j-1)} - v_{g(2j)})$ and $\sum_j(v_{g(2j-1)} - v_{g(2j)})$ are odd if $G = U(n)$, $\sum_j r_j(\pm v_{g(j)})$ and $\sum_j(\pm v_{g(j)})$ are odd for any arrangement of $\pm$ signs if $G = Sp(n)$, and

$$\sum_j r_j(\pm v_{g(2j-1)} - (\pm v_{g(2j)})) \text{ and } \sum_j(\pm v_{g(2j-1)} - (\pm v_{g(2j)}))$$

are odd for any arrangement of $\pm$ signs, requiring only that the number of minus signs be even if $G = SO(2n)$, if $G = SO(2n$ or $2n+1)$. As the r_j are odd, the first number has the same residue mod 2 as the second, so we may ignore the first one in each case. Similarly, we may ignore signs. We now see that the condition for $\Gamma \cap \mathrm{ad}(G)K = 1$ is that, for every $g \in S(n)$, $\sum_{j=1}^{2s} v_{g(j)}$ is odd if $G \neq Sp(n)$; $\sum_{j=1}^{s} v_{g(j)}$ is odd if $G = Sp(n)$. This is independent of a and of the r_j, and depends only on n, s and q. It happens if and only if $n = 2s$ and q is odd if $G \neq Sp(n)$; q is odd or $q = n$ if $G = Sp(n)$.

The theorem is now proved except for $G = SU(n)$. Suppose $G = SU(n) \subset U(n)$ and let K' be a closed connected subgroup of rank $n-1$ in $U(n)$ such that $K = K' \cap SU(n)$. If Γ is a subgroup of $SU(n)$, $\Gamma \cap \mathrm{ad}(SU(n))K = \Gamma \cap \mathrm{ad}(U(n))K'$, so we are done because a generalized quaternionic subgroup of $U(n)$ can be assumed, for purposes of checking $\Gamma \cap \mathrm{ad}(U(n))K' = 1$, to be the image of a sum of S_{r_j} with r_j odd, and hence can be assumed to lie in $SU(n)$. *QED.*

The situation with the **Spin** groups is more complicated because of the relative abundance of faithful **Spin**-representations of the generalized quaternionic groups.

LEMMA 9.1. *Let K be a closed connected subgroup of rank $n-1>0$ in $G = \mathbf{Spin}(2n$ or $2n+1)$ and let -1 be the element of order 2 in the kernel of the natural projection $f\colon \mathbf{Spin} \to SO$. Then K is an even subgroup of G if and only if it contains -1.*

Proof. K contains -1 if and only if -1 lies on a 1-parameter subgroup of K, as K is connected. Let Y be a 1-parameter subgroup of K. As -1 is central in G we may assume that $(f\,Y)(t) = \exp(\sum_s t a_s X_s)$, where the exponential is taken in $SO(2n$ or $2n+1)$ and $\{X_1,\cdots,X_n\}$ is our canonical integral basis for $SO(2n$ or $2n+1)$. We may assume Y normalized so that $(f\cdot Y)(t) = I$ if and only if t is an integer; the a_s are then integers and -1 lies on Y if and only if $-1 = Y(1)$. $Y(1) = -1$ if and only if the number of odd a_j is odd.

Let $V = \{v_1,\cdots,v_n\}$ be the canonical parameter of K in G, i.e., the canonical parameter of $f(K)$ in SO. If K is odd, $\sum_s v_s a_s = 0$ implies that

an even number of a_s are odd and $-1 \neq Y(1)$. If K is even, we can assume v_1 even and v_2 odd; we construct a conjugate X of a 1-parameter subgroup of K which contains -1 by $(f\cdot X)(t) = \exp(t(v_2X_1 - v_1X_2))$. *QED.*

We mention an interesting consequence of Lemma 9.1:

COROLLARY 9.1. *Let K be a closed connected subgroup of rank $n-1>0$ in $G = \boldsymbol{SO}(2n$ or $2n+1)$. Then G/K is simply connected if and only if K is an even subgroup of G; $\pi_1(G/K) = Z_2$ if K is odd.*

Proof. The universal covering of G/K is $\boldsymbol{Spin}(2n$ or $2n+1)/K'$, where K' is the identity component of $f^{-1}(K)$. *QED.*

THEOREM 9.2. *Let K be a closed connected subgroup of rank $n-1>0$ in $G = \boldsymbol{Spin}(2n$ or $2n+1)$, $V = \{v_1,\cdots,v_n\}$ the canonical parameter of K in G, and q the number of v_j which are odd.*

Suppose K is even. Then G has a generalized quaternionic subgroup Γ such that $\Gamma \cap \mathrm{ad}(G)K = 1$ if and only if n is even (say $n = 2s$) and q is odd, and any such Γ is $\mathrm{ad}(G)$-conjugate to the image of a faithful G-representation F' of a $\boldsymbol{Q}2^a$, where $F' = T'_{r_1} + \cdots + T'_{r_s}$ for some choice of odd integers r_j. If $n = 2s$, q is odd, $a \geqq 3$, $\{r_1,\cdots,r_s\}$ are odd integers and

$$\Gamma = (T'_{r_1} + \cdots + T'_{r_s})(\boldsymbol{Q}2^{a'}),$$

then $\Gamma \cap \mathrm{ad}(G)K = 1$. If n is odd or q is even, and Σ is a finite subgroup of G such that $\Sigma \cap \mathrm{ad}(G)K = 1$, then every Sylow subgroup of Σ is cyclic.

Suppose K is odd. Given $a \geqq 3$, G has a subgroup Γ isomorphic to $\boldsymbol{Q}2^a$ such that $\Gamma \cap \mathrm{ad}(G)K = 1$, -1 is the element of order 2 in Γ, $f(\Gamma)$ is a dihedral 2-subgroup of $\boldsymbol{SO}(2n$ or $2n+1) = G'$ such that $f(\Gamma) \cap \mathrm{ad}(G')f(K) = 1$ and $f(\Gamma)$ is $\mathrm{ad}(G')$-conjugate to the image of a non-faithful G'-representation F of $\boldsymbol{Q}2^a$ which is a sum of representations of the forms $S_{2r} + S_{4s}$, r odd, and $S_{2t} + S$, t odd.

Proof. $f\colon G = \boldsymbol{Spin} \to \boldsymbol{SO} = G'$ being the natural projection, -1 is the element of order 2 in ker. f and $K' = f(K)$. Suppose K is even, so $-1 \in K$; given a subgroup Γ of G, $\Gamma \cap \mathrm{ad}(G)K = 1$ if and only if $f(\Gamma) \cap \mathrm{ad}(G')K' = 1$ and $-1 \notin \Gamma$. A generalized quaternionic subgroup Γ of G not containing -1 is $\mathrm{ad}(G)$-conjugate to the image of a faithful G-representation $F' = T'_{r_1} + \cdots + T'_{r_s} + t'$, t' not faithful, of $\boldsymbol{Q}2^a$. $f\cdot F'$ is a faithful G'-representation $F = T_{r_1} + \cdots + T_{r_s} + t$, t not faithful, of $\boldsymbol{Q}2^a$. In the proof of Theorem 9.1 we saw that $f(\Gamma) \cap \mathrm{ad}(G')K' = 1$ if and only if $n = 2s$ and q is odd, and this is independent of the choices of a and of the odd integers r_j.

If $n=2s$, then t, and hence t', is trivial because F represents by matrices of determinant 1.

Now suppose K is odd. Given $a \geqq 3$ we will construct a subgroup Γ' of G', isomorphic to the dihedral group $\boldsymbol{D}2^{a-1}$ of 2^{a-1} elements, such that $\Gamma' \cap \mathrm{ad}(G')K' = 1$ and $\Gamma = f^{-1}(\Gamma')$ is isomorphic to $\boldsymbol{Q}2^a$. Then -1 will be the element of order 2 in Γ. Given $\gamma \in \Gamma \cap \mathrm{ad}(G)K$, $f(\gamma) = 1$; as $-1 \notin K$ it follows that $\gamma = 1$, so $\Gamma \cap \mathrm{ad}(G)K = 1$.

Let Γ' be a dihedral 2-subgroup of G'. As $\boldsymbol{D}2^{a-1}$ is the quotient of $\boldsymbol{Q}2^a$ by the subgroup generated by $B^2 = A^{2^{a-2}}$, we may view Γ' as the image of a representation $F = S_{2r_1} + \cdots + S_{2r_u} + S_{4t_1} + \cdots + S_{4t_v} + s$, r_j odd and s a sum of $\boldsymbol{O}(1)$-representations, of $\boldsymbol{Q}2^a$. Let $2w'$ and w be the multiplicities of the eigenvalue -1 of $s(A)$ and $s(B)$; the eigenvalue -1 of $F(B)$ has multiplicity $u+v+w=2x$. K' is odd because K is odd; as in the proof of Theorem 9.1 the arithmetic criterion shows that $\{F(A)\} \cap \mathrm{ad}(G')K' = 1$ if and only if u is odd when $a > 3$, if and only if $u + w'$ is odd when $a = 3$. It also shows that $\{F(B)\} \cap \mathrm{ad}(G')K' = 1$ if and only if x is odd.

The representations F which lift to ***Spin*** are of the form

$$(*) \qquad \begin{aligned} F = {} & (S_{2r_1} + S_{4s_1}) + \cdots + (S_{2r_p} + S_{4s_p}) \\ & + (S_{2t_1} + S) + \cdots + (S_{2t_q} + S) \end{aligned}$$

with r_j, t_j odd and where $S\colon A \to 1$, $B \to -1 \in \boldsymbol{O}(1)$. In this case, every element of Γ' is conjugate to a power of $F(A)$ or $F(B)$ and we have $u = p + q = x$, $w' = 0$; it follows that $\Gamma' \cap \mathrm{ad}(G')K' = 1$ if and only if $p+q$ is odd. rank. $K > 0$ implies $n > 1$, so we can find non-negative integers p and q with $p+q$ odd and $4p + 3q \leqq 2n$, hence a G'-representation F of the form $(*)$ with $p+q$ odd. *QED.*

V. 2. Subgroups of the unitary group which have all Sylow subgroups cyclic and act freely. Suppose that K is a closed connected subgroup of $\boldsymbol{U}(n)$ of rank $n-1$ and Γ is a finite subgroup of $\boldsymbol{U}(n)$ with all Sylow subgroups cyclic. Γ is conjugate to the image of a faithful representation of an abstract finite group Σ with all Sylow subgroups cyclic, and we can replace Γ by that conjugate. We will apply our arithmetic criterion (Theorem 6) to see whether $\Gamma \cap \mathrm{ad}(\boldsymbol{U}(n))K = 1$. This is of considerable interest if n is odd or the number of odd elements of the canonical parameter of K in $\boldsymbol{U}(n)$ is even. for then every finite subgroup B of $\boldsymbol{U}(n)$ such that $B \cap \mathrm{ad}(\boldsymbol{U}(n))K = 1$ has every Sylow subgroup cyclic.

Let N be the order of Σ. We represent Σ by generators and relations:

$A^m = B^n = 1$, $BAB^{-1} = A^r$, $0 < m$, $mn = N$, $((r-1)n, m) = 1$ and $r^n \equiv 1 \pmod{m}$. If Σ is cyclic, $m = 1$ and this becomes $B^N = 1$.

If Γ is cyclic of order t and has a generator γ with eigenvalues

$$\exp(\pi i r_1/t), \cdots, \exp(2\pi i r_n/t),$$

and $V = \{v_1, \cdots, v_n\}$ is the canonical parameter of K in $\boldsymbol{U}(n)$, a direct application of the arithmetic criterion shows that $\Gamma \cap \mathrm{ad}(\boldsymbol{U}(n))K = 1$ if and only if $\sum_s r_s v_{g(s)}$ is prime to t for every element g of the permutation group $\boldsymbol{S}(n)$. We will, then, ignore this case and henceforth assume that Σ is not cyclic.

In the notation of § IV. 3, we can assume that Γ is the image of the faithful representation $F = \sum_{j=1}^{a} \sum_{p=1}^{b_j} f_{jk_{jp}}$ of Σ in $\boldsymbol{U}(q)$.

Given an integer u and a divisor m_j of m, we define $u^{(j)} = (u, d_j)$, $0 \leqq u_j < d_j$ *and* $u_j \equiv u \pmod{d_j}$, $d_j^{(u)} \cdot u^{(j)} = d_j$, and

$$r^{(u)} = 1 + r^{u_j} + r^{2u_j} + \cdots + r^{(d_j^{(u)}-1)u_j}.$$

Given a second integer v, we define $h(u,v)$ to be the order of $B^u A^v$ in Σ. A calculation shows that $f_{jk_{jp}}(B^u A^v)$ has eigenvalues

$$\exp(2\pi i[(h(u,v)/nm_j)(k_{jp}um_j + u^{(j)}m_j n_j' e + n_j' v u^{(j)} r^{(u)} r^t)]/h(u,v))$$

for $0 \leqq e < d_j^{(u)}$ and $0 \leqq t < u^{(j)}$. If u is prime to d_j, a calculation shows that this means that $f_{jk_{jp}}(B^u A^v)$ has eigenvalues $\exp(2\pi i[k_{jp}u - en_j']/n)$ for $0 \leqq e < d_j$, hence is $\mathrm{ad}(\boldsymbol{U}(d_j))$-conjugate to $f_{jk_{jp}}(B^u)$. If $d \mid u$, another calculation shows that the eigenvalues of $f_{jk_{jp}}(B^u A^v)$ can be written, on setting $u = wd$ so $u/n = w/n'$, as

$$\exp(2\pi i[k_{jp}wm + (mn'/m_j)vr^t]/mn') \text{ for } 0 \leqq t < d_j.$$

Now set $N_{jpe} = \sum_{o<j} d_o b_o + (p-1)d_j + e + 1$, and, given an integer u, set $N_{jpet}(u) = \sum_{o<j} d_c b_o + (p-1)d_j + eu^{(j)} + t + 1$. With this notation, an application of the arithmetic criterion now yields:

THEOREM 10. *Let K be a closed connected subgroup of rank $q-1$ in the unitary group $\boldsymbol{U}(q)$, $V = \{v_1, \cdots, v_q\}$ the canonical parameter of K in $\boldsymbol{U}(q)$, $\boldsymbol{S}(q)$ the group of all permutations on $\{1, 2, \cdots, q\}$ and Γ a finite subgroup of $\boldsymbol{U}(q)$ which is the image of a faithful representation $F = \sum_{j=1}^{a} \sum_{p=1}^{b_j} f_{jk_{jp}}$ of the abstract finite non-cyclic group Σ with all Sylow subgroups cyclic. Then $\Gamma \cap \mathrm{ad}(\boldsymbol{U}(q))K = 1$ if and only if for every $g \in \boldsymbol{S}(q)$ we have:*

1. $\sum_{j=1}^{a}\sum_{p=1}^{b_j}\sum_{e=0}^{d_j-1} v_{g(N_{jpe})}\cdot(k_{jp}+en_j')$ *is prime to* n.

2. $\sum_{j=1}^{a}\sum_{p=1}^{b_j}\sum_{e=0}^{d_j-1} v_{g(N_{jpe})}\cdot(mk_{jp}+(m/m_j)n'r^e)$ *is prime to* mn'.

3. Given integers u and v with $1\leqq u<n$, $1\leqq v<m$ *and* $1<(u,d)<d$,
$$\sum_{j=1}^{a}\sum_{p=1}^{b_j}\sum_{e=0}^{d_j^{(u)}-1}\sum_{t=0}^{u^{(j)}-1} v_{g(N_{jpet}(u))}\cdot(h(u,v)/nm_j)$$
$$\cdot(k_{jp}um_j+u^{(j)}m_jn_j'e+n_j'vu^{(j)}r^{(u)}r^t)\not\equiv 0 \pmod{h(u,v)}.$$

To adapt these formulae to the other classical groups, we proceed as follows:

$\boldsymbol{SU}(q)$. K must have rank $q-2$ and Γ must lie in $\boldsymbol{SU}(q)$. Formulae (1, 2, 3; Theorem 10) remain unchanged.

$\boldsymbol{Sp}(q)$. $\Gamma\subset\boldsymbol{Sp}(q)\subset\boldsymbol{U}(2q)$, $\boldsymbol{S}(q)$ must be replaced by the group $\boldsymbol{S}'(q)$ of all signed permutations on V, and, for each formula of Theorem 10, the numbers following the v's fall into two sets, one of which is the negative of the other, and only one must be summed.

$\boldsymbol{SO}(2q+1)$. $\Gamma\subset\boldsymbol{SO}(2q+1)\subset\boldsymbol{U}(2q+1)$ and we proceed as for $\boldsymbol{Sp}(q)$.

$\boldsymbol{SO}(2q)$. $\Gamma\subset\boldsymbol{SO}(2q)\subset\boldsymbol{U}(2q)$, $\boldsymbol{S}(q)$ must be replaced by the group $\boldsymbol{S}''(q)$ of all signed permutations on V which involve an even number of changes of sign, and we proceed as for $\boldsymbol{Sp}(q)$.

$\boldsymbol{Spin}(2q$ or $2q+1)$. We proceed as for $\boldsymbol{SO}(2q$ or $2q+1)$.

Recall that if $G=\boldsymbol{U}(q)$, $\boldsymbol{SU}(q)$, $\boldsymbol{Sp}(q)$, $\boldsymbol{SO}(2q$ or $2q+1)$, or $\boldsymbol{Spin}(2q$ or $2q+1)$, and if $K=\boldsymbol{U}(q-1)$, $\boldsymbol{SU}(q-1)$, $\boldsymbol{Sp}(q-1)$, $\boldsymbol{SO}(2q-1$ or $2q-2)$, or $\boldsymbol{Spin}(2q-1$ or $2q-2)$, respectively, imbedded in the usual way, the canonical parameter of K in G is $\{1,0,\cdots,0\}$. With this in mind, we can use Theorem 10 to generalize some rather nice theorems of H. Zassenhaus [16] and G. Vincent [14]:

COROLLARY 10.1. *Let G be a classical group $\boldsymbol{U}(q)$, $\boldsymbol{SU}(q)$, $\boldsymbol{Sp}(q)$, $\boldsymbol{SO}(2q$ or $2q+1)$ or $\boldsymbol{Spin}(2q$ or $2q+1)$ and let K be a closed connected subgroup such that* rank. G — rank. $K=1$ *and the canonical parameter of K in G is $\{1,0,0,\cdots,0\}$. Let Γ be a finite subgroup of G with $\Gamma\cap\mathrm{ad}(G)K=1$, such that the order of Γ is either the product of two primes or is prime to $2q$. Then Γ is cyclic.*

Proof. Suppose first that the order of Γ is prime to $2q$. As every abelian subgroup of Γ is cyclic and Γ has odd order, every Sylow subgroup of Γ is cyclic. Formula 2 of Theorem 10 now says that $mk_{jp}+(m/m_j)n'r^e$ is prime to mn', hence to m, so $m_j=m$ and consequently each $d_j=d$. This implies that d divides both q and the order of Γ, which are relatively prime, so $d=1$. But $d=1$ ifplies $r=1$ and thus that Γ is cyclic.

Suppose Γ has order mn with m and n prime, and that Γ is not cyclic. As Γ is not abelian, $m\neq n$. It follows that every Sylow subgroup of Γ is cyclic, so we look at Theorem 10, which, by switching m and n if necessary, is directly applicable. $d\,|\,n$ and n is prime, so $d=1$ or $d=n$. As $d=1$ implies that Γ is cyclic, $d=n$. Formula 2 of Theorem 10 then shows $m_j=m$, so $n_j'=n'=1$. Formula 1 of Theorem 10 then says that $k_{jp}+e$ is prime to n for $0\leqq e<n$, which is impossible. *QED.*

In addition to providing known information on spheres, Corollary 10.1 tells us something about the Grassmann manifolds $\boldsymbol{SO}(2q)/\boldsymbol{SO}(2q-2)$, $\boldsymbol{SO}(2q+1)/\boldsymbol{SO}(2q-1)$ and $\boldsymbol{SO}(2q+1)/\boldsymbol{SO}(2q-2)$. The formulae of Theorem 10 can yield all sorts of information by placing special conditions on the canonical parameter.

Chapter VI. Manifolds with non-zero Euler characteristic.

After stating that we would for the most part concentrate on the case $\operatorname{rank}.G-\operatorname{rank}.K\leqq 1$, we devoted our attention primarily to the case $\operatorname{rank}.G-\operatorname{rank}.K=1$. In this chapter, we will prove a theorem about the case where $\operatorname{rank}.G=\operatorname{rank}.K$. First recall the well-known fact ([10], p. 15) that a coset space G/K of a compact connected Lie group G by a closed subgroup K has Euler characteristic $\chi(G/K)\geqq 0$, and that $\chi(G/K)>0$ if and only if $\operatorname{rank}.G=\operatorname{rank}.K$. We will prove:

THEOREM 11. *Let M be a compact connected Riemannian homogeneous manifold with Euler characteristic $\chi(M)\neq 0$. Then there are only a finite number, up to isometry, of Riemannian manifolds which admit M as a Riemannian covering manifold.*

Remark. If M' admits a Riemannian covering of multiplicity n by M, we have $\chi(M)=n\cdot\chi(M')$. As $\chi(M')$ must be an integer, it is clear, intuitively speaking, that one can go down only a finite number of steps from M. The theorem says, then, that there are only a finite number of steps from M. The theorem says, then, that there are only a finite number of "directions" in which one can go down. These various "directions" will

be seen to correspond roughly to the subgroups of the finite group G/G_0, where G_0 is the identity component of the group G of isometries of M.

Proof. We will first show that we may assume M simply connected, so that we will only have to consider normal coverings, i.e., coverings which are effectuated by the group of deck transformations. Let K_0 be the intersection of an isotropy subgroup K of the group G of isometries of M with the identity component G_0 of G, so $M = G_0/K_0$. K_0 contains a maximal torus of G_0 but contains no nontrivial normal subgroup of G_0; it follows that G_0 is centerless, hence semisimple, and thus has finite fundamental group. The homotopy sequence of the fibring $G_0 \to G_0/K_0 = M$ then shows that M has finite fundamental group, so the universal Riemannian covering manifold M'' is compact. We will be done if we show that only a finite number, up to isometry, of Riemannian manifolds admit a Riemannian covering by M'', so we may assume M simply connected.

We now need only show that there are only a finite number of properly discontinuous subgroups of G which give mutually non-isometric quotient manifolds of M. As conjugate subgroups of G give isometric quotients, we need only show that there are only a finite number of mutually non-conjugate properly discontinuous subgroups of G. As $\operatorname{rank}.G = \operatorname{rank}.K$, $\operatorname{ad}(G)K$ contains G_0 so a properly discontinuous subgroup of G meets G_0 only at 1 and is thus isomorphic to a subgroup of the finite group G/G_0. The proof of Theorem 11 is thus reduced to:

LEMMA 11.1 (MOSTOW). *Let G be a compact Lie group and Γ a finite group. Then G contains only a finite number of conjugacy classes of isomorphs of Γ.*

Proof. Suppose the contrary and let $\{\Gamma_n\}$ be a sequence of mutually non-conjugate isomorphs of Γ which lie in G. We can assume that $\Gamma_n = \{\gamma_{1n}, \gamma_{2n}, \cdots, \gamma_{kn}\}$ ordered so that $\gamma_{jn} \to \gamma_{jm}$ is an isomorphism $\Gamma_n \to \Gamma_m$ for every m and n. As Γ is finite and G compact, we can assume that each sequence $\{\gamma_{jn}\}_n$ converges, $\{\gamma_{jn}\} \to \gamma_j$. It is clear that $\Sigma = \{\gamma_1, \cdots, \gamma_k\}$ is a subgroup of G, although we don't yet know that the γ_j are all distinct. A theorem of D. Montgomery and L. Zippin ([9], p. 216) says that Σ has a neighborhood U such that every subgroup of G in U is $\operatorname{ad}(G)$-conjugate to a subgroup of Σ. As the Γ_n eventually lie in U, this contradicts their mutual non-conjugacy. *QED.*

The proof of Theorem 11 also furnishes a proof of:

THEOREM 11′. *Let M be a compact connected Riemannian homogeneous*

manifold. If M has finite fundamental group, there are only a finite number, up to isometry, of Riemannian manifolds with a given fundamental group which admit M as a Riemannian covering manifold. In any given case, there are only a finite number, up to isometry, of Riemannian manifolds which admit a normal Riemannian covering by M with a given group of deck transformations.

THE UNIVERSITY OF CHICAGO.

REFERENCES.

[1] A. Borel, "Fixed point theorems for elementary abelian groups, I," *Seminar on Transformation Groups*, notes, The Institute for Advanced Study, Princeton, 1959.

[2] ———, "Commutative subgroups and torsion in compact Lie groups," (to appear).

[3] ——— and J, de Siebenthal, "Les sous-groupes fermés de rang maximum des groupes de Lie clos," *Commentarii Mathematici Helvetici*, vol. 23 (1949), pp. 200-221.

[4] C. Chevalley, *Theory of Lie groups*, vol. I, Princeton, 1946.

[5] P. E. Conner, "On the action of a finite group on $S^n \times S^n$," *Annals of Mathematics* (2), vol. 66 (1957), pp. 586-588.

[6] A. Heller, "A note on spaces with operators," *Illinois Journal of Mathematics*, vol. 3 (1959), pp. 98-99.

[7] A. G. Kurosh, *The theory of groups*, vol. I, New York, 1955.

[8] J. Milnor, "Groups which act on S^n without fixed points," *American Journal of Mathematics*, vol. 79 (1957), pp. 623-630.

[9] D. Montgomery and L. Zippin, *Topological transformation groups*, New York, 1955.

[10] H. Samelson, "On curvature and characteristic of homogeneous spaces," *Michigan Mathematical Journal*, vol. 5 (1958), pp. 13-18.

[11] P. A. Smith, "Permutable periodic transformations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 30 (1944), pp. 105-108.

[12] N. Steenrod, *The topology of fibre bundles*, Princeton, 1951.

[13] M. Suzuki, "On finite groups with cyclic Sylow subgroups for all odd primes," *American Journal of Mathematics*, vol. 77 (1955), pp. 657-691.

[14] G. Vincent, "Les groupes linéaires finis sans points fixes," *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 117-171.

[15] B. L. van der Waerden, *Modern Algebra*, vol. II, New York, 1950.

[16] H. Zassenhaus, "Über endliche Fastkörper," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, bd. 11 (1935), pp. 187-220.

[17] ———, *The theory of groups*, second edition, New York, 1958.

FUNDAMENTAL GROUPS OF COMPACT SOLVMANIFOLDS.*

By Louis Auslander.[1]

Introduction. Let S be a connected, simply connected solvable Lie group and let G be a closed subgroup of S. Then the homogeneous space S/G is called a solvmanifold. In the special case where the solvable Lie group is nilpotent, the homogeneous space is called a nilmanifold. The study of nilmanifolds was initiated by Malcev, [1], who succeeded in giving a very detailed account of the structure of such spaces. Recently, much effort has been put into extending the Malcev results to the class of solvmanifolds. The major contributions in this direction have been obtained by Mostow [2] and Wang [4]. However, there still remained open at the completion of this work the following two questions.

Problem 1. Give an algebraic characterization of the fundamental groups of *compact* solvmanifolds.

Problem 2. Give an algebraic characterization of the discrete uniform subgroups of solvable Lie groups.

Now clearly every discrete uniform subgroup is a fundamental group of a compact solvmanifold, but an example of Wang [4] shows that the converse is not true. The purpose of this paper is to prove the following theorem:

THEOREM I. *Every fundamental group of a solvmanifold is the fundamental group of a compact solvmanifold.*

Since an algebraic characterization of the fundamental groups of solvmanifolds is contained in [4], this gives a complete solution to problem 1. Problem 2 still remains open and, in the author's opinion is of considerable interest.

* Received March 1, 1959; revised October 1, 1959.

[1] This paper was in preparation during part of the time the author was being supported under a National Science Foundation Grant.

Part I.

1. Recapitulation. Let Γ be the fundamental group of a solvmanifold. Then Γ can be imbedded as a discrete uniform subgroup of a simply connected solvable Lie group S, where S has at most a finite number of components (see [4]). Further, from the construction of S, it follows that the identity component of S, which we will denote by S_0, contains the commutator subgroup of S. But by [3], $S = F \cdot S_0$, where F is a finite group and the dot denotes the semi-direct product. Hence F is an abelian group. Now $\Gamma_0 = \Gamma \cap S_0$ is a discrete uniform subgroup of S_0 and Γ_0 is of finite index in Γ. If N is the maximal normal analytic nilpotent subgroup of S_0, then, again by the construction of S, F induces the identity transformation on S_0/N. Further, $N_0 = N \cap \Gamma$ is a discrete uniform subgroup of N and Γ_0/N_0 is a free abelian group on, we may assume, s generators. Again by the construction given in [4], if N_0 and $\theta_1, \cdots, \theta_s$ generate Γ, then there exist positive integers λ_i, $i = 1, \cdots, s$, such that $(\theta_i)^{\lambda_i}$ lie on one parameter subgroups in S_0.

2. Existence theorem. Let $\mathcal{S}_0$ be any simply connected, connected solvable Lie group for which there exists a homomorphism h of $\mathcal{S}_0$ onto S_0; then $h^{-1}(\Gamma_0)$ is a uniform subgroup of $\mathcal{S}_0$ and $\mathcal{S}_0/h^{-1}(\Gamma_0)$ is homeomorphic to S_0/Γ_0 (see [2]). Further, if $\mathfrak{N}_0$ is the maximal normal analytic nilpotent subgroup of $\mathcal{S}_0$, then the kernel K of h is contained in $\mathfrak{N}_0$ and $h^{-1}(\Gamma_0) \cap \mathfrak{N}_0$ is a uniform subgroup of $\mathfrak{N}_0$ ([2], prop. 1, p. 22).

Definition. Let F be a group of automorphisms of S_0. We will say that F can be lifted to $\mathcal{S}_0$ if there exists a group F^*, F^* isomorphic to F, of automorphisms of $\mathcal{S}_0$ which map K onto itself and induce the automorphism group F on S_0. If for S_0, F there exists $\mathcal{S}_0$ and F^*, as above, we will say that $\mathcal{S}_0$, F^* is an enlargement of S_0, F.

THEOREM 1. *Let Γ be the fundamental group of a solvmanifold and let $\Gamma \subset S$, where S satisfies the conditions of* § 1. *Assume that there exists an enlargement $\mathcal{S}_0$, F^* of S_0, F. Then if F^* can be imbedded in a torus group T of automorphisms of $\mathcal{S}_0$, such that T acts trivially on $\mathcal{S}_0/\mathfrak{N}_0$, Γ is the fundamental group of a* compact *solvmanifold.*

Proof. Consider $\mathcal{S} = F^* \cdot \mathcal{S}_0$. Then h can be extended to a homomorphism h_* of $\mathcal{S}$ onto S if we define h_* by

$$h_*(f^*, s) = (i(f^*), h(s)),$$

where $f^* \in F^*$, $i(f^*)$ is the image of f^* in F, $s \in \mathcal{S}_0$ and $h(s) \in S$. There

the kernel of h_* equals the kernel of h or K and hence the kernel of h_* is connected, simply connected and contained in $\mathfrak{N}_0$, where $\mathfrak{N}_0$ is the maximal normal analytic nilpotent subgroup of $\mathcal{S}_0$. Consider in $h_*^{-1}(\Gamma)$ pre-images $\eta_1, \cdots, \eta_s$ of $\theta_1, \cdots, {}_s$ in Γ. Since $h_*^{-1}(\Gamma)/h_*^{-1}(\Gamma) \cap \mathfrak{N}_0$ is free abelian on s generators, $h^{-1}(\Gamma) \cap \mathfrak{N}_0$ and $\eta_1, \cdots, \eta_s$ generate $h_*^{-1}(\Gamma)$.

We will use the symbol $L(\ )$ to denote the Lie algebra of the Lie group in the parentheses. Now let $x_i \in L(S_0)$ be such that $\exp(x_i) = \theta_i^{\lambda_i}$. Since h maps $\mathcal{S}_0$ onto S_0, there exists $X_i \in L(\mathcal{S}_0)$ such that X_i maps onto x_i by the induced mapping of $L(\mathcal{S}_0)$ onto $L(S_0)$. Since $h(\exp(X_i)) = \theta_i^{\lambda_i}$, we must have that $\exp(X_i) \equiv \eta_i^{\lambda_i} \bmod K$. Now T is a compact group, T acts trivially on $L(\mathcal{S}_0/\mathfrak{N}_0)$, and $X_i \notin L(\mathfrak{N}_0)$. Hence we can find $Y_i \in L(\mathcal{S}_0)$ such that Y_i is invariant under T and $Y_i \equiv X_i \bmod L(\mathfrak{N}_0)$. Let $\eta_i = (f_i, s_i)$. Since $\mathfrak{N}_0$ is a characteristic subgroup in $\mathcal{S}_0$, it is normal in $\mathcal{S}$ and we can consider $U: S \to \mathcal{S}/\mathfrak{N}_0$. Then $\mathcal{S}/\mathfrak{N}_0 = F^* + V^s$, where V^s is an s dimensional vector space and $+$ denotes the direct sum. Let $\exp((1/\lambda_i)Y_i) = (0, r_i)$. Let $U(\eta_i) = (f_i, \xi_i)$. Then $U(\eta_i^{\lambda_i}) = (f_i^{\lambda_i}, \lambda_i\xi_i) = (0, \lambda_i\xi_i)$. But $U(\exp(\xi_i Y_i)) = U(0, r_i^{\lambda_i}) = (0, \lambda_i\xi_i)$. Hence $U(\exp((1/\lambda_i)Y_i)) = (0, \xi_i)$. This shows that $f^*_i((1/\lambda_i)Y_i) \equiv \eta_i \bmod \mathfrak{N}_0$.

Now let $S^{\#} = T \cdot \mathcal{S}_0$, where the dot denotes the semi-direct product. Consider the subalgebra A of $L(\mathcal{S}^{\#})$ spanned by $L(\mathfrak{N}_0)$ and $(t_i + (1/\lambda_i)Y_i)$, where $t_i \in L(T)$ such that $\exp(t_i) = f^*_i$. Let $\mathcal{S}^*$ be the connected subgroup whose Lie algebra is A. We wish to show that $\mathcal{S}^*$ is simply connected and closed in $\mathcal{S}^{\#}$. Now $\mathcal{S}^* \supset \mathfrak{N}_0$ and $\mathcal{S}^*/\mathfrak{N}_0$ is a vector subgroup of the abelian Lie group $T \cdot \mathcal{S}_0/\mathfrak{N}_0$. Since both $\mathcal{S}^*/\mathfrak{N}_0$ and $\mathfrak{N}_0$ are closed and simply connected, the same is true of $\mathcal{S}^*$. Hence $p^{-1}(\Gamma) \cap \mathcal{S}^*$ is a closed subgroup of $\mathcal{S}^*$, since each is a closed subset of $\mathcal{S}^{\#}$. But $\mathcal{S}^* \supset p^{-1}(\Gamma)$. For $\mathcal{S}^* \supset \mathfrak{N}_0$ and $\mathcal{S}^*$ contains $\exp(t_i + (1/\lambda_i)Y_i)$. But

$$\exp(t_i + (1/\lambda_i)Y_i) = \exp(t_i)\exp((1/\lambda_i)Y_i)$$

since $[t_i, Y_i] = 0$. Hence $\exp(t_i + (1/\lambda_i)Y_i) \equiv \eta_i \bmod \mathfrak{N}_0$. Hence $\mathcal{S}^* \supset \eta_i$, $i = 1, \cdots, s$. This shows that $\mathcal{S}^* \supset h_*^{-1}(\Gamma)$. It remains to verify that $\mathcal{S}^*/h_*^{-1}(\Gamma)$ is compact. This follows easily from the fact that $h_*^{-1}(\Gamma) \cap \mathfrak{N}_0$ is uniform in $\mathfrak{N}_0$ and the fact that in $\mathcal{S}^*/\mathfrak{N}_0$ the images of $\eta_1, \cdots, \eta_s$ generate a discrete uniform subgroup. This proves that $\mathcal{S}^*/h_*^{-1}(\Gamma)$ can be represented as a fiber bundle with compact fiber and base. Hence $\mathcal{S}^*/h_*^{-1}(\Gamma)$ is compact and our theorem is proved.

3. A decomposition of $L(S_0)$. Let $L(S_0)$ denote the Lie algebra of S_0. Then, since F is a finite abelian group, and F acts trivially on $L(S_0/N_0)$,

where $L(N_0)$ is invariant under all automorphisms of $L(S_0)$, there is a direct sum decomposition of $L(S_0)$

$$L(S_0) = L(N_0) + C,$$

where F acts trivially on C. However, since F is abelian, $L(N_0)$ can be divided into the direct sum of invariant two dimensional spaces $V^2{}_\alpha$ $\alpha = 1, \cdots, r$, and one dimensional spaces $V^1{}_\beta$, $\beta = 1, \cdots, t$. We will consider each $V^2{}_\alpha$ as oriented and given an inner product with respect to which F acts as orthogonal transformations. Now, we may define an equivalence relations on the objects $V^2{}_\alpha$ and divide them into equivalence classes $\mathcal{U}^2{}_i$, $i = 1, \cdots, r'$, as follows:

$V^2{}_\alpha$ is said to be equivalent to $V^2{}_\alpha$ or $V^2{}_\alpha \sim V^2{}_{\alpha'}$ if either

1. For all $f \in F$, f induces the same positive angle of rotation in $V^2{}_\alpha$ and $V^2{}_{\alpha'}$,

or

2. For all $f \in F$, f induces the negative of the angle of rotation in $V^2{}_\alpha$ as in $V^2{}_{\alpha'}$.

It is clear that in each equivalence class $\mathcal{U}^2{}_i$ each member can be so oriented that condition 1 will always hold. We will for the rest of this paper assume that this has been done.

LEMMA 1. *Let $V^2{}_\alpha \in \mathcal{U}^2{}_i$, let $X \in C$ and let $[X, V^2{}_\alpha]$ denote the linear mapping of the linear space spanned by $V^2{}_\alpha$ under bracket with X. Then either*

a) *$[X, V^2{}_\alpha]$ is a linear transformation of $V^2{}_\alpha$ onto $V^2{}_{\alpha'} \in \mathcal{U}^2{}_i$ (i, the same as in the hypothesis)* or b) *$[X, V^2{}_\alpha]$ maps $V^2{}_\alpha$ into zero.*

Proof. Let (Y, Z) be an oriented orthonormal basis for $V^2{}_\alpha$. Then consider the linear space spanned by $[X, Y]$ and $[X, Z]$. Then there exists $f \in F$ such that

$$f(Y) = \cos\theta Y + \sin\theta Z,$$
$$f(Z) = -\sin\theta Y + \cos\theta Z,$$

where $|\cos\theta| < 1$ and $0 \leqq \theta \leqq 2\pi$. But

$$f[X, Y] = \cos\theta[X, Y] + \sin\theta[X, Z],$$
$$f[X, Z] = -\sin\theta[X, Y] + \cos\theta[X, Z].$$

Now if $[X, Y] = 0$, then $[X, Z]$ must be zero. For otherwise

$$f[X, Z] = \cos\theta[X, Z]$$

where $|\cos\theta| \neq 1$. Since F is a finite group, this is impossible. An analogous argument shows that if $[X, Z] = 0$, then $[X, Y] = 0$, also. Hence the image of $V^2{}_\alpha$ under $[X, V^2{}_\alpha]$ is a 2 dimensional invariant space or it is the null space. But if the image $V^2{}_\alpha$ under $[X, V^2{}_\alpha]$ is a 2 dimensional invariant space this space must also be in $\mathfrak{U}^2{}_i$. This proves the lemma.

COROLLARY. *Let $V^2{}_\alpha$ and $V^3{}_\beta$ be given and let h_1 and h_2 be two linear mappings of $V^2{}_\alpha$ onto $V^2{}_\beta$ induced by a succession of bracketing with elements $X \in C$. Then if $V^2{}_\alpha$ is oriented, the orientations of $V^2{}_\beta$ induced by h_1 and h_2 are compatible.*

Proof. Let (Y, Z) be oriented bases for $V^3{}_\alpha$ and let the linear transformation f operate on $V^2{}_\beta$ with $0 < |\cos\theta| < 1$. Then relative to the bases $(h_1(Y), h_1(Z))$ and $(h_2(Y), h_2(Z))$ the linear transformation f has exactly the same matrix. This implies that the orientation of $V^3{}_\beta$ induced by h_1 and h_2 are compatible.

By the above corollary we may orient coherently all elements of $\mathfrak{U}^2{}_i$ with the property that one gets mapped onto another by an element of C.

Since each $V^2{}_\alpha \in \mathfrak{U}^2{}_i$ has an inner product and an orientation assigned to it, we may consider the one parameter group of orientation preserving rotations. Let T_t denote this one parameter group of orientation preserving rotations of the vector spaces in $\mathfrak{U}^2{}_i$. Then on $V^2{}_\alpha$, T_t will be defined explicitly by:

$$T_t(Y_\alpha) = \cos 2\pi t Y_\alpha + \sin 2\pi t Z_\alpha,$$

$$T_t(Z_\alpha) = -\sin 2\pi t Y_\alpha + \cos 2\pi t Z_\alpha,$$

$0 \leqq t \leqq 1$. Note that for any other compatibly oriented orthonormal basis for $V^2{}_\alpha$ the form of the equations expressing the action of T_t would be identical.

LEMMA 2. *Let $V^2{}_\alpha \in \mathfrak{U}^2{}_i$, $X \in C$ and $t \in T_t$. Then*

$$t[X, Y] = [X, tY]$$

for $Y \in V^2{}_\alpha$.

Proof. If the image of $V^2{}_\alpha$ under bracket with X is zero, the lemma is trivially true. Hence assume this is not the case and that (Y_α, Z_α) is an oriented orthonormal basis of $V^2{}_\alpha$. Then $([X, Y_\alpha], [X, Z_\alpha])$ will be oriented orthonormal basis of the image of $V^2{}_\alpha$ under $[X, V^2{}_\alpha]$. Hence T_t induces

the same linear transformations relative to the basis $([X, Y_\alpha], [X, Z_\alpha])$ of the image of $V^2{}_\alpha$ as it does relative to the basis (Y_α, Z_α) for $V^2{}_\alpha$. This proves the lemma.

4. Construction of enlargements. It is now our purpose to show that the hypothesis of Theorem 1 can always be satisfied. We will restrict ourselves to the obvious Lie algebra formulation of the hypothesis and show that this can always be satisfied. Since any automorphism group of a Lie algebra induces an isomorphic automorphism group of any associated simply connected Lie group and every algebra has a simply connected, connected associated Lie group, once we have shown that the hypothesis can be satisfied for Lie algebras, this will prove that this is also the case for Lie groups. The proof that the hypothesis of the existence theorem can always be satisfied will be arrived at in two stages. Step one will consist in showing that there exists an *infinite* dimensional Lie algebra with certain properties. Step two will consist in finding a finite homomorphic image of the infinite Lie algebra which will satisfy the hypothesis of Theorem 1.

Let $w_1, \cdots, w_k$ be any finite set of symbols. Then we may form the free Lie algebra over the reals, $L(k)$, consisting essentially of the module with the non-associative words in $w_1, \cdots, w_k$ as basis divided by the ideal of the "proper" relations. We will now make this more precise.

Let W be the set of non-associative words in the letters $w_1, \cdots, w_k$ form the module $M(W)$ of linear combinations over the reals of the elements of W. Now define the bilinear mapping of $M(W) \times M(W) \to M(W)$ which extends the mapping which maps the pair of words (u, w) into the word $u(w)$. In this algebra, consider the ideal I generated by the relations of the form

1. $x \cdot x = 0$

2. $(x \cdot (y \cdot z)) + (y \cdot (z \cdot x)) + (z \cdot (x \cdot y)) = 0$

for $x, y, z \in M(W)$. Then $L(k)$ is defined as $M(W)/I$. We will call $L(k)$ the fee Lie algebra on k generators. $L(k)$ is clearly an infinite dimensional vector space over the reals.

LEMMA 3. *Let $V(W)$ be the sublinear space of $M(W)$ spanned by the elements $w_1, \cdots, w_k$. Then every non-singular linear mapping A of $V(W)$ onto itself is uniquely extendable to an automorphism A^* of $L(k)$ onto itself.*

Proof. Let $W_1, \cdots, W_k$ and $w_1, \cdots, w_k$ be two sets of k symbols. Then

the free Lie algebra generated by each of these sets of symbols are clearly isomorphic under the mapping which sends every word in the lower case letters into the corresponding word with capital letters. Now let $A(w_i), \cdots, A(w_k)$ be a new basis for $V(W)$. Then call $A(w_i) = W_i$. Then the isomorphism described above determines an automorphism of $L(k)$ onto itself.

LEMMA 4. *Any k dimensional Lie algebra is the homomorphic image of $L(k)$.*

This lemma is obvious.

5. **Infinite enlargements.** In $L(S_0)$ we have already discussed the sets $V^2{}_\alpha$, $V^1{}_\beta$ and C. Let I be the linear subspace of $L(S_0)$ spanned by vectors of $V^1{}_\beta$ which are invariant under all elements of F. Let V be the subspace spanned by the remaining elements of $V^1{}_\beta$. Then $V^2{}_\alpha$, I, V and C form a direct sum decomposition of $L(S_0)$, where $V^2{}_\alpha$, I, V form a direct sum decomposition of $L(N_0)$. Let (Y_α, Z_α) be an oriented basis for $V^2{}_\alpha$, as discussed in § 3, let $X_1^*, \cdots, X_s^*$ be a basis for I. Let U be a finite dimensional vector space isomorphic to V with a fixed isomorphism ϕ chosen in advance. Let $U_1, \cdots, U_k$ be the images of $W_1, \cdots, W_k$ under ϕ. Let L be the free Lie algebra generated by the symbols Y, Z, X^*, W, X and U under the appropriate ranges of subscripts. Then we can define a homomorphism of L onto $L(S_0)$ by mapping all U into zero and every other letter into the corresponding vector in $L(S_0)$. Call this homomorphism h.

In order to define an enlargement, we have to define an automorphism group F^* acting on L. But our Lemma 3, showed this was equivalent to defining a group of automorphisms of the linear space L spanned by the symbols Y, Z, etc. Now there is a natural isomorphism of the subspace spanned by all the symbols except U onto $L(S_0)$. Use this isomorphism to define F^* on this subspace of L. Complete the definition of F^* as follows: Define

$$f^*U_i = \operatorname{sign}(f^*, W_i)\, U_i,$$

where $\operatorname{sign}(f^*, W_i)$ is defined by the equation

$$fW_i = \operatorname{sign}(f^*, W_i)\, W_i.$$

This determines the group F^* of automorphism of L. F^* preserves the kernel of the homomorphism h and induces F on $L(S_0)$, since

$$\begin{array}{ccc} L & \xrightarrow{f^*} & L \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ L(S_0) & \xrightarrow{f} & L(S_0) \end{array}$$

is a commutative diagram.

LEMMA 5. *The automorphism group F^* can be imbedded in a torus group T^* of automorphism of L.*

Proof. Define the circle group T_i^* for L: Define it on the vector space spanned by the letters Y_α, Z_α, for $V^2_\alpha \in \mathfrak{U}^2_i$ as in § 3, define it to act trivially on all other letters. This defines T_i^* on L by Lemma 3. Define a circle group T on L as follows:

$$T(W_A) = \cos 2\pi t W_A + \sin 2\pi t U_A,$$
$$T(U_A) = -\sin 2\pi t W_A + \cos 2\pi t U_A,$$

$A = 1, \cdots, k$, $0 \leqq t \leqq 1$ and T acts trivially on all other symbols. Consider the torus group $\mathfrak{T}$ generated by the circle groups T_i^* and T^*. $\mathfrak{T}$ clearly contains F^* as a subgroup. This proves the lemma.

6. Finite enlargements. We are now faced with the problem of finding an $\mathcal{S}$ "between" L and S_0 which is finite dimensional and such that the kernel of the homomorphism which maps L onto S is invariant under $\mathfrak{T}$. This will then induce a torus group T on $\mathcal{S}$ and the pair $\mathcal{S}$, T will satisfy the hypothesis of Theorem 1. This will then prove Theorem I stated in the introduction.

Let $L(N_0)$ be ξ step nilpotent. Consider in L the ideal I^* generated by the following words.

1. All words of length $\xi + 1$ not containing any letters X.
2. If in $L(S_0)$ $[X_i, X_j] = \sum a_{ijs} X^*_s$, then the word $[X_i, X_j] - \sum a_{ijs} X^*_s$ in L.
3. If in $L(S_0)$ $[X_i, W_A] = \sum a_{iAB} W_B$, then the word $[X_i, W_A] - \sum a_{iAB} W_B$ in L.
4. If in $L(S_0)$ $[X_i, W_A] = \sum a_{iAB} W_B$, then the word $[X_i, U_A] - \sum a_{iAB} U_B$ in L.

5. If in $L(S_0)$ $[X_i, Y_\alpha] = Y_{\alpha'}$, then the word $[X_i, Y_\alpha] - Y_{\alpha'}$ in L.

6. If in $L(S_0)$ $[X_i, Z_\alpha] = Z_{\alpha'}$, then the word $[X_i, Z_\alpha] - Z_{\alpha'}$ in L.

LEMMA 6. *$\mathcal{S} = L/I^*$ is a finite dimensional solvable Lie algebra and h induces a homomorphism of $\mathcal{S}$ into $L(S_0)$. Further, I^* is invariant under $\mathcal{T}$ in L.*

Proof. To see that $\mathcal{S}$ is finite dimensional and solvable, we note that $[\mathcal{S}, \mathcal{S}]$ consists of the images of words involving no X. But since all words not invloving X of length greater than $\xi + 1$ map into zero in $\mathcal{S}$, we see that $[\mathcal{S}, \mathcal{S}]$ is a finite dimensional nilpotent Lie algebra. Further, $\mathcal{S}/[\mathcal{S}, \mathcal{S}]$ is isomorphic to C and hence of finite dimension. This shows that $\mathcal{S}$ is a finite dimensional solvable Lie algebra. To see that h induces a homomorphism of $\mathcal{S}$ onto $L(S_0)$, we have only to note that I^* is certainly in the kernel of h.

It remains to show that I^* is invariant under $\mathcal{T}$. We have only to show that I^* is invariant under the one parameter groups T^*_i and T^*. Clearly, if we see that the vector space spanned by the generators of I^* is invariant under these groups, then I^* is invariant under these groups. But by Lemma 2 this vector space is invariant under the groups T^*_i. To verify that this is also true for T^* is trivial. This proves the lemma.

This completes the proof of Theorem I.

REFERENCES.

[1] A. I. Malcev, "On a class of homogeneous spaces," *American Mathematical Society, Translations*, No. 39.

[2] G. D. Mostow, "Factor spaces of solvable groups," *Annals of Mathematics*, vol. 60 (1954), pp. 1-27.

[3] ———, "Self-adjoint group," *ibid.*, vol. 62 (1955), pp. 44-55.

[4] H. C. Wang, "Discrete subgroups of solvable Lie groups," *ibid.*, vol. 64 (1956), pp. 1-19.

COMPACT TRANSFORMATION GROUPS OF S^n WITH AN $(n-1)$-DIMENSIONAL ORBIT.*

By Hsien-Chung Wang.[1]

1. Introduction. The compact and connected transformation groups of an n-sphere S^n with an $(n-1)$-dimensional orbit were first studied by Montgomery and Zippin [10]. When G has a fixed point, they determined the orbits and reduced this problem to the problem of compact, transitive transformation groups of homotopy spheres. Recently, under weaker conditions on the dimension of the orbits but assuming differentiability, Montgomery-Samelson-Yang show that G must act on S^n orthogonally. Later, Poncet [17] proved this without differentiability condition. When G has no fixed point, the situation is different and some results have been obtained by Mostert [18] and Nagano [19]. It is the aim of the present paper to classify these transformation groups when the action is differentiable. The main results can be stated as follows:

"*Let G be a connected, compact and effective differentiable transformation group of S^n with an $(n-1)$-dimensional orbit. Suppose that either n is even and different from 4, or is odd and greater than* 31. *Then G acts on S^n orthogonally, and up to an equivalence in the sense of transformation groups, G falls into five main classes. If a subgroup of the rotation group $O(n+1)$ is regarded as a transformation group of the unit sphere in R^{n+1}, then these classes can be described as follows:* (i) *G is a subgroup of $O(n+1)$ leaving invariant and transitive on the unit sphere S^{n-1} in a hyperplane of R^{n+1}.* (ii) *Let G', G'' be subgroups of $O(p+1)$, $O(q+1)$ transitive on the unit spheres S^p, S^q in R^{p+1}, R^{q+1} respectively, where $p>0$, $q>0$, $p+q=n-1$, and denote by $G'\times G''$ the direct product in the sense of linear groups.*[2] *G is a subgroup of $O(n+1)$ containing $G'\times G''$ as a normal sub-*

* Received September 24, 1959.

[1] This work was partially supported by the National Science Foundation.

[2] Let G' be a group of automorphisms of R^a, and G'' a group of automorphisms of R^b. By the direct product $G'\times G''$ in the sense of linear group, we mean the group $G'\times G''$ acting on the direct sum $R^a\oplus R^b$ in the natural manner.

group. (iii) *G is the tensor product $Sp(r)\otimes Sp(2)$ over the quaternions*[3] *regarded as a subgroup of $O(8r)$, $n=8r-1$, where $Sp(m)$ denotes the group of all $(m\times m)$ quaternion symplectic matrices.* (iv) *G is either the tensor product $SU(r)\otimes SU(2)$ or $U(r)\otimes U(2)$ regarded as subgroups of $O(4r)$, $n=4r-1$, where $U(m)$ denotes the unitary group and $SU(m)$ the unimodular unitary group.* (v) *G is the tensor product $O(p+2)\otimes O(2)$ regarded as a subgroup of $O(2p+4)$, $n=2p+3$."*

We note that in Cases (i) and (ii), the classification of G is reduced to the classification of compact transitive transformation groups of spheres which is known.

Our proof consists of two parts. In the first part, we establish some general properties of the orbits and isotropic subgroups. In fact, from the results of Montgomery-Zippin, all the $(n-1)$-dimensional orbits X are homeomorphic, and besides, there are two orbits X_1, X_2 of lower dimension. Let L, L_1, L_2 be the isotropic subgroups corresponding to X, X_1, X_2 respectively. Then $X=G/L$, $X_1=G/L_1$, $X_2=G/L_2$. We can choose the isotropic subgroups so that $L\subset L_1$, $L\subset L_2$. From these two inclusions, two projections $f_1\colon X\to X_1$, $f_2\colon X\to X_2$ are defined. Let C_i be the mapping cylinder associated to f_i $(i=1,2)$. Then S^n can be written as the union $C_1\cup C_2$, where $C_1\cap C_2=X$. By using this decomposition, it is proved that f_1 and f_2 have the following properties:[4]

(A) X is a sphere bundle over X_i with f_i as the projection $(i=1,2)$.

(B) Let f_i^* be the homomorphism of the cohomology with any coefficient induced by f_i $(i=1,2)$. Then f_i^* is injective, and, for $0<m<n-1$,

$$f_1^*(H^m(X_1))+f_2^*(H^m(X_2))=H^m(X),\qquad f_1^*(H^m(X_1))\cap f_2^*(H^m(X_2))=0.$$

(C) Let $f_i^\#$ be the homomorphism of the fundamental groups induced by f_i. Then $f_1^\#(\ker f_2^\#)=\pi_1(X_1)$, $f_2^\#(\ker f_1^\#)=\pi_1(X_2)$.

If the differentiability is not assumed, (B) and (C) are still valid, but we can only prove that X is a homology sphere bundle over X_i.

We note that (B) and (C) reflect the global properties of S^n while (A)

[3] Let Q^r be the r-dimensional right quaternion vector space, and Q^2 the left quaternion plane. In the suitable manner, $Sp(r)$ and $Sp(2)$ can be regarded as the groups of symplectic transformations of Q^r and Q^2 respectively. Thus $Sp(r)\times Sp(2)$ acts naturally on the tensor product $Q^r\otimes Q^2$ which is a real vector space R^{8r}, and the action is real-linear. This group of automorphisms of R^{8r} is called the tensor product $Sp(r)\otimes Sp(2)$.

[4] Properties (A) and (B) have been proved independently by Mostert and Nagano.

is concerned with certain local properties. In fact, we have a partial converse. Let X, X_1, X_2 be compact manifolds and X orientable. Suppose $f_i\colon X \to X_i$ to be a fibre map with fibre F_i of positive dimension $(i = 1, 2)$. Denote by C_i the mapping cylinder associated with f_i, and by Σ the union of C_1 and C_2 with their lower bases X identified. Then condition (B) implies that Σ is a homology sphere while (C) implies that Σ is simply-connected. Since X is a manifold, Σ is locally euclidean at any point in $\Sigma - (X_1 \cup X_2)$. At points in X_i, the local homology depends on the homology of the fibre F_i.

The second part consists in the determination of the transformation group G. Taking account of the properties (A), (B), (C) and using known results on the real cohomology of homogeneous spaces, we can classify the system $\{G, L, L_1, L_2\}$ when $n > 31$. On the other hand, it can be proved that the action of G is completely determined by the triple $\{L, L_1, L_2\}$ of subgroups. In this way, the transformation groups $\{G, S^n\}$ are determined.

The author wishes to express his sincere thanks to Dr. G. E. Bredon and Professor H. Samelson for their valuable comments.

2. Mapping cylinder. For later use, we shall establish in this section a property about mapping cylinders. Let X, Y be two spaces, and $f\colon X \to Y$ a continuous surjective map. Suppose that any set V in Y is open if and only if $f^{-1}(V)$ is open in X. We define the mapping cylinder C associated with f to be the space obtained from the product $X \times I$ $(I = \{t\colon 0 \leqq t \leqq 1\})$ by identifying the points $(x, 1)$ and $(x', 1)$ when $f(x) = f(x')$. In the natural manner, C can be written as the union $X \cup (X \times (0,1)) \cup Y$, where $(0,1) = \{t\colon 0 < t < 1\}$, and X, Y retain their own topologies. We shall always take X and Y as closed subsets of C in this sense. X and Y will be called the lower and upper bases of C respectively. The map $r\colon C \to C$ defined by $r(x) = f(x)$, $r(y) = y$, $r((x,t)) = f(x)$, $x \in X$, $y \in Y$, $0 < t < 1$, is a retraction and is homotopic to the identity.

Let B be a subset of Y and $A = f^{-1}(B)$. Then the set

$$C_A = A \cup (A \times (0,1)) \cup B$$

is the mapping cyliner associated to $f|_A\colon A \to B$. We shall call C_A the sub-cylinder of C over A.

From the definition, it follows directly that

(2.1) *Let C be the mapping cylinder associated to a map $f\colon X \to Y$, and U a neighborhood in C of a point y of Y. If $f^{-1}(y)$ is compact, then there exists a real number η, $0 < \eta < 1$, such that $f^{-1}(y) \times [1-\eta, 1) \subset U$, where $[1-\eta, 1) = \{t\colon 1-\eta \leqq t < 1\}$.*

Now we can prove the following lemma:

(2.2) *Let $f\colon X \to Y$ be a fibre map with fibre F, where Y, F are compact manifolds, not necessarily connected, and C the mapping cylinder associated to f. If $C - X$ is a differentiable manifold and the imbedding of Y in $C - X$ is differentiable, then F is a homotopy sphere.*

Proof. Let $\dim Y = m$, $\dim F = r$. Then $\dim X = m + r$, and $\dim C = m + r + 1$. Take an open m-cell b^m in Y, and set $A = f^{-1}(b^m)$. Then A is the product bundle over b^m with fibre F. Now consider the subcylinder C_A of C over A. We know that $C_A = A \cup (A \times (0,1)) \cup b^m$, and that $C_A - A$ is an $(m + r + 1)$-dimensional differentiable manifold containing b^m as a differentiable submanifold. Let y be a point of b^m. There exists a neighborhood U of y in C_A such that the difference set $U - b^m$ is homeomorphic with the product $R^{m+1} \times S^r$, where S^r denotes the r-sphere and R^{m+1} the $(m+1)$-dimensional euclidean space. By (2.1), we can find η, $0 < \eta < 1$, with the property that $F \times [1 - \eta, 1) \subset U$, where $F = f^{-1}(y)$ is a fibre. Evidently

$$F \times [1 - \eta, 1) \subset U - b^m \subset C_A - b^m.$$

Since $C_A - b^m$ is homeomorphic with $F \times b^m \times [0,1)$ and b^m is contractible, $F \times [1 - \eta, 1)$ is a deformation retract of $C_A - b^m$, and hence the inclusion maps

$$\pi_i(F \times [1 - \eta, 1)) \to \pi_i(C_A - b^m), \qquad i = 0, 1, 2, \cdots$$

are injective. From the fact that $U - b^m$ is homeomorphic with $R^{m+1} \times S^r$, it follows that $\pi_i(F) = \pi_i(F \times [1 - \eta, 1)) = 0$ for $0 \leqq i \leqq r - 1$, and that $\pi_r(F)$ has no torsion. By assumption, F is a compact manifold, and therefore F must be a homotopy sphere of dimension r.

3. Transformation groups of S^n with an $(n-1)$-dimensional orbit. Let G be a connected and compact group acting on S^n almost effectively with an $(n-1)$-dimensional orbit. Montgomery-Zippin [10, 11] proved that G is a Lie group, and that, with two exceptions, all the orbits under G are homeomorphic to one another. Moreover, there exists an arc A in S^n such that (i) each orbit intersects A in exactly one point, (ii) the isotropic subgroups of G at inner points of the arc A are all identical, and (iii) the orbits at the end points of A have dimension less than $n - 1$. Now choose an inner point a of the arc A. Let a_1, a_2 be the end points of A, and $X = G(a)$, $X_1 = G(a_1)$, $X_2 = G(a_2)$ be the respective orbits. Denote by L, L_1, L_2 the isotropic subgroups of G at a, a_1, a_2 respectively. Then $L \subset L_1$, $L \subset L_2$,

$X = G/L$, $X_1 = G/L_1$, $X_2 = G/L_2$. Since G is a Lie group, X, X_1, X_2 are manifolds. The map $f_i\colon X \to X_i$ defined by $f_i(gL) = gL_i$, $g \in G$, is a fibre map with L_i/L as the fibre $(i = 1, 2)$.

Let us write A as the union of two arcs A_1 and A_2, where $A_1 = \widehat{\dot{a}_1 a}$, $A_2 = \widehat{a_2 a}$. For each i, the orbit $G(A_i)$ of A_i under G is evidently the mapping cylinder C_i associated to the map $f_i\colon X \to X_i$. Thus we can write

$$S^n = C_1 \cup C_2, \qquad C_1 \cap C_2 = X. \tag{3.1}$$

From now on, let us assume that G acts on S^n differentiably. It follows that X_i is a submanifold imbedded in S^n, and hence in $C_i - X$, in the differentiable manner. (2.2) tells us then that L_i/L must be a homotopy sphere. From a theorem of Borel[5] [2], we know that L_i/L is actually a sphere $(i = 1, 2)$.

For each i, let $k_i\colon X \to C_i$ be the injection map, and $r_i\colon C_i \to X_i$ be the natural retraction of the mapping cylinder onto its upper base. Evidently

$$f_1 = r_1 k_1, \qquad f_2 = r_2 k_2. \tag{3.2}$$

Moreover, it is well-known that the homomorphism

$$r_i^*\colon H^*(X_i) \to H^*(C_i) \tag{3.3}$$

of the cohomology rings induced by the retraction r_i is bijective. Now we apply the Vietoris-Mayer sequence to (3.1). It follows at once that the homomorphism $H^s(C_1) + H^s(C_2) \to H^s(X)$ defined by $(v_1, v_2) \to k_1^*(v_1) - k_2^*(v_2)$, $v_1 \in H^s(C_1)$, $v_2 \in H^s(C_2)$, is bijective when $0 < s < n-1$. This fact together with (3.2) and (3.3) tells us that the homomorphism $H^s(X_1) + H^s(X_2) \to H^s(X)$ given by $(u_1, u_2) \to f_1^*(u_1) - f_2^*(u_2)$, $u_1 \in H^s(X_1)$, $u_2 \in H^s(X_2)$, is bijective for $0 < s < n-1$.

Summarizing the above results, we have

(3.4) *Let G be a connected and compact group acting differentiably on S^n with an $(n-1)$ dimensional orbit. Denote by X an $(n-1)$-dimensional orbit, and by X_1, X_2 the two particular orbits. There are natural maps $f_1\colon X \to X_1$, $f_2\colon X \to X_2$ which are sphere fibre maps. Let $f_i^*\colon H^*(X_i) \to H^*(X)$ denote the homomorphism of the cohomology rings with arbitrary coefficient ring $(i = 1, 2)$. Then f_i^* is injective, and, for $0 < s < n-1$,*

$$f_1^*(H^s(X_1)) + f_2^*(H^s(X_2)) = H^s(X), \qquad f^*(H^s(X_1)) \cap f_2^*(H^s(X_2)) = 0.$$

[5] When n is even, this follows from known results on homogeneous spaces with non-vanishing Euler characteristic. For arbitrary n, Borel's method [1] of determining the compact transitive transformation groups of spheres gives a definite way to prove this theorem, however, detail was not given. Matsushima has also shown this when n is odd and not too small. Poncet [17] gave the proof and provided the detail for the exceptional cases.

Now suppose X_i to be orientable. Since X is orientable, $\{X, X_i, f_i\}$ is an orientable sphere bundle [7], and hence we have the Gysin-Chern-Spanier sequence with integer coefficients

$$H^0(X_i) \to H^{m+1}(X_i) \xrightarrow{f_i^*} H^{m_i+1}(X) \to \cdots,$$

where $m_i = \dim X - \dim X_i = \dim L_i - \dim L$. Since f_i^* is injective, the characteristic class of the sphere bundle always vanishes. Therefore,

(3.5) *If X_i is orientable, then the sphere bundle $\{X, X_i, f_i\}$ has zero characteristic class (integer coefficients). Moveover, X and $X_i \times S^{m_i}$ have the same Poincaré polynomial.*

Since we always have the Gysin-Chern-Spanier sequence with integers mod 2 as the coefficient field, then

(3.6) *X and $X_i \times S^{m_i}$ have the same Poincaré polynomial if the coefficient field is of characteristic 2.*

4. Fundamental group of orbits. In this section, some properties of the fundamental group of the orbits will be established by using a theorem of van Kampen. We first digress a little and prove a proposition on free product of groups.

(4.1) *Let A, B_1, B_2 be finitely generated abstract groups, and $h_i : A \to B_i$ be surjective homomorphisms with kernel K_i $(i = 1, 2)$. Let N be the least normal subgroup of the free product $B_1 \square B_2$ containing the set $\{h_1(a)h_2(a^{-1}) : a \in A\}$ of elements in $B_1 \square B_2$. Then $A/K_1K_2 \approx (B_1 \square B_2)/N$.*

Proof. For any m symbols $x_1, \cdots, x_m$ and any set $\{P_\alpha\}$ of words in m variables, let us denote by $[x_1, \cdots, x_m : P_\alpha(x)]$ the group generated by $x_1, \cdots, x_m$ with $\{P_\alpha(x_1, \cdots, x_m) = 1\}$ as the set of defining relations. Since A is finitely generated, and h_1, h_2 are surjective, we can write

$$(4.2) \quad \begin{aligned} A = [a_1, \cdots, a_m; P_\alpha(a)], \qquad B_1 &= [b_1, \cdots, b_m; P_\alpha(b), Q_\beta(b)], \\ B_2 &= [c_1, \cdots, c_m; P_\alpha(c), R_\gamma(c)], \end{aligned}$$

where $b_i = h_1(a_i)$, $c_i = h_1(a_i)$, and $\{P_\alpha\}$, $\{Q_\beta\}$, $\{R_\gamma\}$ are three sets of words in m variables. It is evident that

$$B_1 \square B_2 = [b_1, \cdots, b_m, c_1, \cdots, c_m; P_\alpha(b), P_\alpha(c), Q_\beta(b), R_\gamma(c)]$$

and that N is the least normal subgroup of $B_1 \square B_2$ containing $b_1c_1^{-1}$, $\cdots, b_mc_m^{-1}$. Therefore

$$\begin{aligned}(B_1 \square B_2)/N \approx [b_1,\cdots,b_m, c_1,\cdots,c_m;\\ P_\alpha(b), P_\alpha(c), Q_\beta(b), R_\gamma(c), b_1c_1^{-1},\cdots,b_mc_m^{-1}]\\ \approx [b_1,\cdots,b_m; P_\alpha(b), Q_\beta(b), R_\gamma(b)]\\ \approx [a_1,\cdots,a_m; P_\alpha(a), Q_\beta(a), R_\gamma(a)].\end{aligned}$$

On the other hand, we see immediately from (4.2) that K_1 and K_2 are, respectively, the least normal subgroups of A containing the sets $\{Q_\beta(a_1,\cdots,a_m)\}$ and $\{R_\gamma(a_1,\cdots,a_m)\}$ of elements. It follows then that $(B_1\square B_2)/N \approx AK_1K_2$ which proves our proposition.

Now we return to the discussion of the transformation group of spheres, and let $G, S^n, X, X_1, X_2, \cdots$ have the same meaning as before.

(4.3) *Let $f_i^\#\colon \pi_1(X)\to\pi_1(X_i)$ be the homomorphisms of the fundamental groups induced by the projection $f_i\colon X\to X_i$, and K_i be the kernel of $f_i^\#$ $(i=1,2)$. Then $\pi_1(X)=K_1\cdot K_2$, and $f_1^\#(K_2)=\pi_1(X_1)$, $f_2^\#(K_1)=\pi_1(X_2)$.*

Proof. We recall that $S^n = C_1\cup C_2$, $C_1\cap C_2 = X$. Let $\pi_1(C_1)\square\pi_1(C_2)$ denote the free product of $\pi_1(C_1)$ and $\pi_2(C_2)$, and N the least normal subgroup of $\pi_1(C_1)\square\pi_1(C_2)$ containing all the elements of the form $k_1^\#(a)k_2^\#(a^{-1})$, $a\in\pi_1(X)$, where $k_i^\#\colon \pi_1(X)\to\pi_1(C_i)$ are the homomorphisms of the fundamental groups induced by the injection map $k_i\colon X\to C_i$ $(i=1,2)$. Since $C_i - X$ is open in S^n and X, C_i are locally contractible, a theorem of van Kampen [12] tells us that $\pi_1(S^n)=\pi_1(C_1\cup C_2)\approx(\pi_1(C_1)\square\pi_1(C_2))/N$. Hence

$$N = \pi_1(C_1)\square\pi_1(C_2). \tag{4.4}$$

Let $r_i\colon C\to X_i$ be the retraction of the mapping cylinder onto its upper base. We know that r_i is homotopic to the identity map of C_i, whence $r_i^\#\colon\pi_1(C_i)\to\pi_1(X_i)$ is bijective $(i=1,2)$. Since $f_i = r_ik_i$, the equality (4.4) tells us that

$$N^* = \pi_1(X_1)\square\pi_1(X_2), \tag{4.5}$$

where N^* is the least normal subgroup of $\pi_1(X_i)\square\pi_1(X_2)$ containing all the elements of the form $f_1^\#(a)f_2^\#(a^{-1})$, $a\in\pi_1(X)$. We know that the projection $f_i\colon X\to X_i$ is a fibre map with the homotopy sphere of positive dimension as the fibre, and therefore $f_i^\#$ is surjective. It follows from (4.1) and (4.5) that $\pi_1(X)=K_1K_2$. Hence

$$\pi_1(X_1)=f_1^\#(\pi_1(X))=f_i(K_1K_2)=f_1^\#(K_2), \pi_1(X_2)=f_2^\#(\pi_1(X))=f_2^\#(K_1K_2) = f_2^\#(K_i).$$

(4.6) *Both* $\pi_1(X_1)$ *and* $\pi_1(X_2)$ *are cyclic, and* $K_1 \cap K_2$ *is the commutator subgroup of* $\pi_1(X)$. *Let* $p = \dim X - \dim X_1$, $q = \dim X - \dim X_2$. *Then* (i) *for* $p > 1$ *and* $q > 1$, *the orbits* X, X_1, X_2 *have trivial fundamental group, and* (ii) *when* $p = 1$ *and* X_1 *orientable, the circle bundle* X *over* X_1 *with the projection* f_1 *is the product bundle and* $\pi_1(X)$ *is abelian.*

Proof. We know that X is a fibre bundle over X_1 with the homotopy sphere S^p as the fibre, and thus we have the following exact sequence

$$\cdots \to \pi_1(S^p) \to \pi_1(X) \xrightarrow{f_1{}^{\#}} \pi_1(X_1) \to \cdots.$$

The kernel K_1 of $f_1{}^{\#}$, being a homomorphic image of $\pi_1(S^p)$, is cyclic. It follows then from (4.3) that $\pi_1(X_2)$ is cyclic. Similarly, we can show that $\pi_1(X_1)$ is cyclic. To see the intersection $K_1 \cap K_2$, let us denote by $f_{i*}\colon H_1(X, Z) \to H_1(X_i, Z)$ the homomorphisms of the 1-dimensional homology groups induced by $f_i\colon X \to X_i$. We have the following commutative diagram:

$$\begin{array}{ccc} \pi_1(X) & \xrightarrow{f^{\#}} & \pi_1(X_1) \times \pi_1(X_2) \\ {\scriptstyle m}\downarrow & & \downarrow{\scriptstyle l} \\ H_1(X, Z) & \xrightarrow{f_*} & H_1(X_1, Z) + H_1(X_2, Z), \end{array}$$

where m, l are the natural homomorphisms of the fundamental groups onto the 1-dimensional integral homology group, and $f^{\#}$, f_* are defined by

$$f^{\#}(a) = (f_1{}^{\#}(a), f_2{}^{\#}(a)), a \in \pi_1(X); \qquad f_*(f_{1*}(x), f_{2*}(x)), x \in H_1(X, Z);$$

the mapping $f_{i*}\colon H_1(X, Z) \to H_1(X_i, Z)$ denoting the homomorphisms induced by f_i. Since $\pi_1(X_1)$, $\pi_1(X_2)$ are abelian, l is bijective, and by (3.4), f_* is also bijective. Therefore, $f^{\#}$ and m have the same kernel. But the kernel of m is the commutator subgroup of $\pi_1(X)$ while the kernel of $f^{\#}$ is the intersection $K_1 \cap K_2$. The first part of our proposition is thus proved.

If $p > 1$, $q > 1$, it follows directly from the homotopy sequences for the bundles $\{X, X_i, f_i\}$ $(i = 1, 2)$ that $K_1 = K_2 = 0$, and hence by (4.3) that

$$\pi_1(X) = \pi_1(X_1) = \pi_1(X_2) = 0.$$

Now suppose that $p = 1$ and X_1 is orientable. Then X is an orientable circle bundle over X, hence a principal bundle. From (3.5), this bundle has zero characteristic class, and then it must be the product bundle. It follows that $\pi_1(X)$ is isomorphic with the direct product of $\pi_1(X_1)$ and a free cyclic group and therefore is abelian.

(4.7) *Suppose that X_1 is not a point. Then the map $f: X \to X_1 \times X_2$ defined by $f(x) = (f_1(x), f_2(x))$, $x \in X$, is a covering map of X onto $f(X)$.*

Proof. Let us write X, X_1, X_2 in the form of coset spaces G/L, G/L_1, G/L_2. Then we see immediately that $f(gL) = f(g'L)$ if and only if $g^{-1}g' \in L_1 \cap L_2$, where $g, g' \in G$. Thus $f: X \to f(X)$ is a fibre map with $(L_1 \cap L_2)/L$ as the fibre. To prove our proposition (4.7), it suffices to prove that $(L_1 \cap L_2)/L$ is finite. We know that $f_1: G/L \to G/L_1$ is a fibre map with the homotopy sphere S^p as the fibre. Choose the particular fibre $F_1 = \{gL: g \in L_1\}$ in G/L. On account of (3.5) and (3.6), the fundamental cycle of F_1 gives a non-zero element y in the p-dimensional homology group $H_p(G/L, A)$ with coefficient group A, where A denotes the free cyclic group or the cyclic group of order two according as G/L_1 is orientable or not. Let $f_{i*}: H_p(G/L, A) \to H_p(G/L_i, A)$ be the homomorphism of the homology groups induced by the projection $f_i: G/L \to G/L_i$ $(i = 1, 2)$, and

$$f_*: H_p(G/L, A) \to H_p(G/L_1, A) + H_p(G/L_2, A)$$

be defined by $f_*(u) = (f_{1*}(u), f_{2*}(u))$, $u \in H_p(G/L, A)$. By our assumption, X_1 is not a point so $p < n - 1$, and then from the dual of (3.4), f_* is injective. Since $f_{1*}(y) = 0$, $f_{2*}(y)$ cannot be zero and therefore $f_2(F_1)$ must have dimension $\geqq p$. On the other hand, we see directly that $f_2(F_1)$ is homeomorphic with the coset space $L_1/(L_1 \cap L_2)$. This tells us that

$$\dim(L_1 \cap L_2) \leqq \dim L_1 - p.$$

But $\dim L = \dim L_1 - p$ and $L \subset L_1 \cap L_2$, whence $\dim(L_1 \cap L_2) = \dim L$. It follows that $(L_1 \cap L_2)/L$ is finite and thus (4.7) is proved.

(4.8) *If $\dim X_1 + \dim X_2 = \dim X$, then f is a homeomorphism of X onto $X_1 \times X_2$. Therefore, both bundles $\{X, X_1, f_1\}$, $\{X, X_2, f_2\}$ are product bundles and X_1, X_2 are spheres.*

Proof. Since $\dim X_2 < \dim X$, X_1 cannot be a point, and therefore $f(X)$ has the same dimension as X. From our hypothesis, $f(X)$ is a closed submanifold of $X_1 \times X_2$ of the same dimension, whence $f(X) = X_1 \times X_2$. Thus X is a covering space of $X_1 \times X_2$. Identifying $\pi_1(X_1 \times X_2)$ with $\pi_1(X_1) \times \pi_1(X_2)$, we know that the homomorphism $f^{\#}: \pi_1(X) \to \pi_1(X_1 \times X_2)$ induced by the map $f: X \to X_1 \times X_2$ is given by $f^{\#}(a) = (f_1^{\#}(a), f_2^{\#}(a))$, $a \in \pi_1(X)$. By (4.3) $f^{\#}$ is surjective, and therefore the covering map $f: X \to X_1 \times X_2$ must be a homeomorphism.

Let $l_i: X_1 \times X_2 \to X_i$ be the projection $(i = 1, 2)$. Then f establishes an

equivalence of the bundle $\{X, X_i, f_i\}$ with the product bundle $\{X_1 \times X_2, X_i, l_i\}$. (4.8) is thus proved.

(4.9) *Let L^0, L_i^0 denote, respectively, the identity components of L, L_i, and $\Pi_i = (L_i^0 \cap L)/L^0$. Then $(L/L^0)/\Pi_i \approx L_i/L_i^0$ is cyclic $(i = 1, 2)$ and $L/L^0 = \Pi_1 \cdot \Pi_2$.*

Proof. Let us consider the following commutative diagram

$$\begin{array}{ccccc} \pi_1(G) & \longrightarrow & \pi_1(X_1) & \xrightarrow{\theta_1} & L_1/L_1^0 \\ & & \uparrow{\scriptstyle f_1^\#} & & \uparrow{\scriptstyle \epsilon_1} \\ \pi_1(G) & \longrightarrow & \pi_1(X) & \xrightarrow{\theta} & L/L^0 \\ & & \downarrow{\scriptstyle f_2^\#} & & \downarrow{\scriptstyle \epsilon_2} \\ \pi_1(G) & \longrightarrow & \pi_1(X_2) & \xrightarrow{\theta_2} & L_2/L_2^0, \end{array}$$

where the horizontal sequences are the homotopy sequences associated to the bundles $X_1 = G/L_1$, $X = G/L$, $X_2 = G/L_2$, and ϵ_1, ϵ_2 are the natural projections. Evidently Π_i is the kernel of ϵ_i, and since θ, θ_1, θ_2 are surjective, we have $(L/L^0)/\Pi_i \approx \epsilon_i(L/L^0) = L_i/L_i^0$. By (4.6), $\pi_1(X_i)$ is cyclic, so L_i/L_i^0 is also cyclic. To see the last part of (4.9), let K_i denote the kernel of $f_i^\#$. Then $\theta(K_i) \subset \Pi_i$ and then from (4.3) it follows that

$$\Pi_1 \cdot \Pi_2 \supset \theta(K_1) \cdot \theta(K_2) = \theta(\pi_1(X)) = L/L^0,$$

which completes the proof of (4.9).

5. The orbits under G when they are all simply connected. In this section and the next, we shall use the known results on the real cohomology algebras of the homgeneous spaces of compact Lie groups established by Samelson, Leray, Koszul and H. Cartan. The notations of H. Cartan [4] will be used throughout. In fact let Q be a compact Lie group. We shall always use $S(Q)$ to denote the algebra of real symmetric multilinear forms over the Lie algebra of Q, and $I(Q)$ to denote the subalgebra of $S(Q)$ consisting of the fixed elements under the adjoint group of Q. For any principal bundle with base B and group Q, there is a characteristic map $\sigma\colon I(Q) \to H^*(B)$, where H^* denotes the real cohomology algebra. The image of σ is called the *characteristic subalgebra* of $H^*(B)$.

Now we consider our transformation group G of S^n with $(n-1)$-dimensional orbits. Let $X, X_1, X_2, L, L_1, L_2, \cdots$ retain their meaning as

before, and let

$$p = \dim X - \dim X_1 = \mathrm{dm}\, L_1 - \dim L,$$
$$q = \dim X - \dim X_2 = \dim L_2 - \dim L.$$

Assume $p > 1$, $q > 1$. It follows from (4.6) that X, X_1, X_2 are simply connected, and hence L, L_1, L_2 are connected. From [4], we have the following commutative diagram

$$\begin{tikzcd}
 & I(L_1) \arrow[r, "\sigma_1"] \arrow[d, "\alpha_1"] & H^*(X_1) \arrow[d, "f_1^*"] \arrow[dr, "h_1^*"] & \\
I(G) \arrow[ur, "\rho_1"] \arrow[r, "\rho"] \arrow[dr, "\rho_2"'] & I(L) \arrow[r, "\sigma"] & H^*(X) \arrow[r, "h^*"] & H^*(G) \\
 & I(L_2) \arrow[r, "\sigma_2"] \arrow[u, "\alpha_2"] & H^*(X_2) \arrow[u, "f_2^*"] \arrow[ur, "h_2^*"'] &
\end{tikzcd}$$

where H^* denotes the real cohomology algebra, ρ, ρ_1, ρ_2, α_1, α_2 the restriction maps, σ, σ_1, σ_2 the characteristic maps, and h^*, h_1^*, h_2^* the homomorphisms induced, respectively, by the natural projections $h\colon G \to X$, $h_1\colon G \to X_1$, $h_2\colon G \to X_2$.

Suppose that both p and q are odd. L_1/L and L_2/L are odd-dimensional spheres and therefore $\alpha_1 : I(L_1) \to I(L)$ and $\alpha_2 : I(L_2) \to I(L)$ are surjective. Let K, K_1, K_2 be, respectively, the characteristic subalgebras of $H^*(X)$, $H^*(X_1)$, $H^*(X_2)$. Then we have, from diagram (5.1),

$$f_1^*(K_1) = f_1^*\sigma_1(I(L_1)) = \sigma\alpha_1(I(L_1)) = \sigma(I(L))$$
$$= K = \sigma\alpha_2(I(L_2)) = f_2^*\sigma_2(I(L_2)) = f_2^*(K_2).$$

On account of (3.4), K, K_1, K_2 can contain only elements of zero degree, and hence h^*, h_1^*, h_2^* are injective. Let $P(G)$, $P_1(G)$, $P_2(G)$ denote, respectively, the linear subspace of $h^*(H^*(X))$, $h_1^*(X^*(X_1))$, $h_2^*(H^*(X_2))$ consisting of the primitive elements of $H^*(G)$. By a theorem of Samelson [13], we have for $i = 1, 2$,

$$H^*(X) \approx h^*(H^*(X)) \approx \Lambda(P(G)),$$
$$H^*(X_i) \approx h_i^*(H^*(X_i)) \approx \Lambda(P_i(G)),$$

and hence

$$\dim H^*(X) = 2^b, \quad \dim H^*(X_i) = 2^{b_i}, \quad b = \dim P(G), \quad b_i = \dim P_i(G),$$

where Λ denotes the exterior algebra, and dimension here is used in the sense

of that of linear spaces over the real field. Evidently, $P_1(G)+P_2(G)\subset P(G)$, and from (3.4), $P_1(G)\cap P_2(G)=0$. Therefore, $\dim H^*(X)\geqq 2^{b_1+b_2}$. On the other hand, we know from (3.5) that $\dim H^*(X)=2\dim H^*(X_i)$, whence $b_1\leqq 1$, $b_2\leqq 1$.

If one of b_1, b_2 is equal to zero, say b_i, then X_1 is reduced to a point on account of its orientability, and so is X_2. Now assume that none of b_1, b_2 is equal to zero. Then $b_1=b_2=1$, or what is the same, there exist positive integers r, s such that $H^*(X_1)\approx H^*(S^r)$, $H^*(X_2)\approx H^*(S^s)$. It follows then from (3.4) that the Poincaré polynomial of X is $1+t^r+t^s+t^{n-1}$. But X, X_1, X_2 are all orientable, and so $r+s=n-1$, $r=\dim X_1$, $s=\dim X_2$. Proposition (4.8) then tells us that X_1 is a q-sphere, X_2 is a p-sphere and X is homeomorphic with the product $X_1\times X_2$.

Thus we have proved

(5.2) *Let $p=\dim X-\dim X_1$, $q=\dim X-\dim X_2$. If $p>1$, $q>1$, and both p and q are odd, then either* (i) *X_1, X_2 are points, or* (ii) *X_1 is a q-sphere, X_2 is a p-sphere and X is homeomorphic with the product $X_1\times X_2$.*

6. The case by which both p and q are even.

(6.1) *Suppose that both p and q are even. Then X, X_1, X_2 have non-vanishing Euler characteristic.*

Proof. Let $\mathfrak{P}(t)$, $\mathfrak{P}_1(t)$, $\mathfrak{P}_2(t)$ be the Poincaré polynomials of X, X_1, X_2 respecetively. From (3.4) and (3.5), we have

$$\mathfrak{P}(t)=\mathfrak{P}_1(t)+\mathfrak{P}_2(t)+t^{n-1},\qquad \mathfrak{P}(t)=(1+t^p)\mathfrak{P}_1(t)=(1+t^q)\mathfrak{P}_2(t).$$

Let $\mathfrak{Q}(t)$ be the polynomial obtained from $\mathfrak{P}(t)$ by omitting all the terms of even degree, and $\mathfrak{Q}_1(t)$, $\mathfrak{Q}_2(t)$ are similarly obtained from $\mathfrak{P}_1(t)$, $\mathfrak{P}_2(t)$. since p and q are even,

$$\mathfrak{Q}(t)=(1+t^p)\mathfrak{Q}_1(t)=(1+t^q)\mathfrak{Q}_2(t),$$

$$\mathfrak{Q}(t)=\begin{cases}\mathfrak{Q}_1(t)+\mathfrak{Q}_2(t), \text{ if } n-1 \text{ is even}\\ \mathfrak{Q}_1(t)+\mathfrak{Q}_2(t)+t^{n-1}, \text{ if } n-1 \text{ is odd.}\end{cases}$$

Suppose that $n-1$ is even. Then we have $t^{p+q}\mathfrak{Q}_1(t)=\mathfrak{Q}_1(t)$ which implies that $\mathfrak{Q}(t)=\mathfrak{Q}_1(t)=\mathfrak{Q}_2(t)=0$. Therefore X, X_1, X_2 have non-vanishing Euler characteristic if $n-1$ is even.

Now suppose that $n-1$ is odd. A direct calculation tells us that $(t^{p+q}-1)\mathfrak{Q}_1(t)=t^{n-1}(1+t^q)$. This is impossible as $t^{n-1}(1+t^q)$ is not divisible by $t^{p+q}-1$. Therefore $n-1$ cannot be odd, and the proof is thus completed.

(6.2) *Suppose that p and q are even. Then L_1 and L_2 generate the entire group G.*

Proof. Let us consider the commutative diagram

$$\begin{tikzcd}
& I(L_1) \arrow[r, "\sigma_1"] \arrow[d, "\alpha_2"] & H^*(X_1) \arrow[d, "f_1^*"] \\
I(G) \arrow[ur, "\rho_1"] \arrow[r, "\rho"] \arrow[dr, "\rho_2"'] & I(L) \arrow[r, "\sigma"] & H^*(X) \\
& I(L_2) \arrow[r, "\sigma_2"] \arrow[u, "\alpha_1"] & H^*(X_2).\arrow[u, "f_2^*"']
\end{tikzcd}$$

On account of (3.4), we have

$$\begin{aligned} H^0(X) &\supset f_1^*\sigma_1(I(L_1)) \cap f_2^*\sigma_2(I(L_2)) \\ &= \sigma\alpha_1(I(L_1)) \cap \sigma\alpha_2(I(L_2)) = \sigma(\alpha_1(I(L_1)) \cap \alpha_2(I(L_2))). \end{aligned}$$

Denote by $I^+(G)$ the ideal of $I(G)$ consisting of elements of positive degree, and by J the ideal of $I(L)$ generated by $\rho(I^+(G))$. Then [4] J is the kernel of σ, and hence

$$\alpha_1(I^+(L_1)) \cap \alpha_2(I^+(L_2)) \subset J.$$

Since X, X_1, X_2 have non-vanishing Euler characteristic, the groups G, L, L_1, L_2 have the same rank. It follows that the maps ρ, ρ_1, ρ_2, α_1, α_2 are injective and the characteristic maps σ, σ_1, σ_2 are surjective. Thus we can consider $I(G)$, $I(L_1)$, $I(L_2)$ simply to be subalgebras of $I(L)$, and neglect the identity maps ρ, ρ_1, ρ_2, α_1, α_2. Then

$$I^+(G) \subset I^+(L_1) \cap I^+(L_2) \subset J = I(L) \cdot I^+(G),$$

and then

$$J = I(L) \cdot I^+(G) \subset I(L) \cdot \{I^+(L_1) \cap I^+(L_2)\} \subset J,$$

or what is the same,

$$I(L) \cdot \{I^+(L_1) \cap I^+(L_2)\} = J = \text{kernel of } \sigma. \tag{6.3}$$

Let Q be the least closed subgroup of G which contains L_1, L_2. Since both L_1 and L_2 are connected, Q is connected. Then $I(G) \subset I(Q) \subset I(L_1) \cap I(L_2)$, whence $J = I(L) \cdot I^+(Q)$. L and Q have the same rank so

$$H^*(X) = \sigma(I(L)) \approx I(L)/J = I(L)/(I(L) \cdot I^+(Q)) \approx H^*(Q/L).$$

Therefore, $\dim X = \dim Q/L$, $\dim G = \dim Q$ and $G = Q$. (6.2) is thus proved.

For further discussions of the orbits, we need the following:

(6.4) *Let U be a compact, connected Lie group, and Q a closed normal subgroup of U. There exists a connected, closed, normal subgroup M of U such that*

$$U = M \cdot Q, \qquad M \cap Q = \text{finite}.$$

This M is unique unless both Q and U/Q have center of positive dimension.

Proof. This can be proved directly by using Lie algebras.

7. Determination of the orbits when p and q are even. Now we are in a position to determine the orbits X, X_1, X_2 when p and q are even. We assume that G acts effectively on S^n and so it acts effectively on X. Let N_1 be the largest normal subgroup of L_1 which is contained in L. This subgroup N_1 can also be regarded as the maximal subgroup of L_1 acting trivially on L_1/L. Thus L_1/N_1 is an effective and transitive transformation of the p-sphere L_1/L with L/N_1 as the isotropic subgroup. Suppose that N_2 is similarly defined.

(7.1) *If p is even and $\neq 6$, then $(L_1/N_1, L/N_1) = (O(p+1), O(p))$. If $p = 6$, then $(L_1/N_1, L/N_1)$ is either $(O(7), O(6))$ or (G_2, A_2), where $O(m)$ denotes the proper orthogonal group of degree m, G_2 the exceptional Lie group of dimension* 16, *and A_2 the unimodular unitary group of rank* 2.

Proof. This is a known result [16].

(7.2) *Suppose that p and q are even and G acts effectively on S^n. Then $N_1 \cap N_2 = e$, and either* (i) *$L = N_1 \cdot N_2$, or* (ii) *$p = q$, N_1 and N_2 are finite, or* (iii) *$p = q = 4$, N_1 and N_2 are locally isomorphic with $O(3)$, and L is locally isomorphic with $O(3) \times O(3) \times O(3)$.*

Proof. Since L_1/N_1 has no center, by (6.4), there exists a unique connected normal subgroup M_1 of L_1 such that $L_1 = M_1 N_1$ and the intersection $M_1 \cap N_1$ is a finite central subgroup of L_1. We know that N_1, N_2 are normal subgroups of L and so $N_1 \cap N_2$ is normal in N_1. Since M_1 and N_1 commute elementwise, it follows that $N_1 \cap N_2$ is a normal subgroup of L_1. Similarly, we can prove that $N_1 \cap N_2$ is a normal subgroup of L_2. On account of (6.2), $N_1 \cap N_2$ must be normal in G. But $N_1 \cap N_2 \subset L$ and G acts effectively on X. It follows that $N_1 \cap N_2 = e$.

To prove the others parts of our proposition, let us assume that $L \neq N_1N_2$. We find it convenient to divide it into two cases and discuss them separately.

Case 1. Either N_1 or N_2 is finite. To be precise, let N_1 be finite. Then L_1 acts almost effectively on the p-sphere L_1/L. Hence L is simple when $p \neq 4$ and is locally isomorphic with $O(4)$ when $p = 4$. In the former alternative, the subgroup N_2, being normal in L, can only be finite or the whole group L. But by our assumption, $L \neq N_2$ and so N_2 must be finite. In the second alternative, N_2 can only be finite or locally isomorphic with $O(3)$. If N_2 is locally isomorphic with $O(3)$, then L/N_2 is locally isomorphic with $O(4)/O(3)$ which contradicts the fact that $(L_2/N_2)/(L/N_2) = L_2/L$ is an even sphere. Hence both N_1 and N_2 are finite in any alternative. It follows that L/N_1 and L/N_2 are locally isomorphic, and then $p = q$ on account of (7.1).

Case 2. Both N_1 and N_2 are infinite. Consider the projection ϕ: $L \to L/N_1$. Then $\phi(N_2)$ is a normal subgroup of L/N_1. Since $N_1 \cap N_2 = e$, $\phi(N_2)$ is isomorphic with N_2, and hence infinite. On the other hand, from our assumption $L \neq N_1N_2$, we know that $\phi(N_2) \neq L/N_1$. It follows that L/N_1 cannot be simple, and then $p = 4$ and $N_2 \approx \phi(N_2)$ is locally isomorphic with $O(3)$. Similarly, we can prove that $q = 4$, and N_1 is locally isomorphic with $O(3)$. By (7.1), $L/N_1 \approx O(4)$ and so L is locally isomorphic with $O(3) \times O(3) \times O(3)$. This completes the proof of (7.2).

(7.3) *Suppose that p, q are even and $L = N_1 \times N_2$. Then either* (i) *G is of rank 2 and L is a two-dimensional torus, or* (ii) *X_1 is a q-sphere and X_2 a p-sphere.*

Proof. From (6.4), there exists a connected normal subgroup M of L_1 such that $L_1 = N_1M$ and $N_1 \cap M$ is a finite central subgroup of L_1. Let V be the identity component of $L \cap M$. Since $L = N_1 \times N_2 \subset L_1$, we see immediately that

$$L = N_1 \cdot V, \qquad N_1 \cap V = \text{finite}.$$

Taking account of the equality $L = N_1 \times N_2$, we know from (6.4) that either $V = N_2$ or both N_1, N_2 have infinite center. Let us discuss these two cases separately.

Case 1. Both N_1 and N_2 have infinite center. We know that L_1/N_1 acts transitively and effectively on the p-sphere L_1/L with L/N_1 as isotropic subgroup. Since $L/N_1 \approx N_2$ has infinite center, p must be 2 and L/N_1 must be the circle group on account of (7.1). Similarly, $q = 2$ and L/N_2 is

the circle group. Hence N_1, N_2 are circle groups, L is the two-dimensional toral group and G of rank 2.

Case 2. Either N_1 or N_2 has a finite center. Then from the above discussions, we know that $N_2 = V$, or what is the same $N_2 \subset M$. Since L and L_1 have the same rank, M and N_2 have the same rank. Therefore N_2 contains the center of M, and hence contains $M \cap N_1$. Then $M \cap N_1 = M \cap N_1 \cap N_2 = e$ which tells us that $L_1 = N_1 \times M$. Similarly, there exists a connected normal subgroup M' of L_2 such that $L_2 = M' \times N_2$ and $N_1 \subset M'$. It is to be noted that we do not know whether M and M' commute.

Now let use use the tensor notation to write down the algebras of the invariant symmetric multilinear forms:

$$I(L) = I(N_1) \otimes I(N_2),\ I(L_1) = I(N_1) \otimes I(M),\ I(L_2) = I(M') \otimes I(N_2).$$

Thus we have the following commutative diagram

$$\begin{array}{ccc} I(L_1) = I(N_1) \otimes I(M) & \xrightarrow{\sigma_1} & H^*(X_1) \\ \downarrow \alpha_1 & & \downarrow f_1^* \\ I(L) = I(N_1) \otimes I(N_2) & \xrightarrow{\sigma} & H^*(X) \\ \uparrow \alpha_2 & & \uparrow f_2^* \\ I(L_2) = I(M') \otimes I(N_2) & \xrightarrow{\sigma_2} & H^*(X_2). \end{array}$$

Since

$$N_2 \subset M,\ N_1 \subset M',\ \operatorname{rank}(M) = \operatorname{rank}(N_2),\ \operatorname{rank}(M') = \operatorname{rank}(N_1),$$

we can consider $I(M)$ to be a subalgebra of $I(N_2)$, and $I(M')$ to be a subalgebra of $I(N_1)$. Therefore,

$$I(L_1) \cap I(L_2) = I(M') \otimes I(M) \subset I(N_1) \otimes I(N_2).$$

Let $J_1 = I(N_1) \cdot I^+(M')$, $J_2 = I(N_2) \cdot I^+(M)$ be, respectively, the ideals generated by $I^+(M')$, $I^+(M)$ in $I(N_1)$, $I(N_2)$. Then

$$I(L) \cdot \{I^+(L_1) \cap I^+(L_2)\} = J_1 \otimes I(N_2) + I(N_1) \otimes J_2.$$

Taking account of (6.3), we obtain then

$$\begin{aligned} H^*(X) \approx I(L)/J &= \{I(N_1) \otimes I(N_2)\}/\{J_1 \otimes I(N_2) + I(N_1) \otimes J_2\} \\ &\approx I(N_1)/J_1 \otimes I(N_2)/J_2. \end{aligned}$$

On the other hand, M/N_2 is homeomorphic with the p-sphere L_1/L, and M'/N_1

with the q-sphere L_2/L. Since both p and q are even, $H^*(M/N_2)$ and $H^*(M'/N_1)$ coincide with their respective characteristic subalgebras, i. e.,

$$H^*(S^p) \approx H^*(M/N_2) \approx I(N_2)/J_2, \qquad H^*(S^q) \approx H^*(M'/N_1) \approx I(N_1)/J_1.$$

It follows then that $H^*(X) \approx H^*(S^p) \otimes H^*(S^q)$, and then by (3.5), $H^*(X_1) \approx H^*(S^q)$, $H^*(X_2) \approx H^*(S^p)$. This tells us that $\dim X_1 + \dim X_2 = \dim X$. From (4.8), we know then that $X_1 = S^q$, $X_2 = S^p$, $X = S^p \times S^q$.

(7.4) *Suppose that p and q are even and G acts effectively on S^n. Let us use $A_m(B_m, C_m, D_m, G_2, F_4, E_6, E_7, E_8)$ to denote a compact group in the Cartan's class $A_m(B_m, C_m, D_m, G_2, F_4, E_6, E_7, E_8)$, where the lower index signifies the rank. Then there are only five possible cases, namely:*

(I) *X_1 is a q-sphere, X_2 is a p-sphere and $X = X_1 \times X_2$;*

(II) *X_1 and X_2 are single points and X is an $(n-1)$-sphere;*

(III) *G is a compact simple Lie group of rank 2 without center, and L is the two-dimensional toral group, and L_1, L_2 are locally isomorphic with the product of a circle group and A_1;*

(IV) *$p = q = 4$, $n = 13$, G is the adjoint group of C_3, L is locally isomorphic with $C_1 \times C_1 \times C_1$, L_1 and L_2 are locally isomorphic with $C_1 \times C_2$, and the orbits X_1, X_2 are quaternion projective planes;*

(V) *$p = q = 8$, $n = 25$, $G = F_4$, $L = D_4$, $L_1 = B_4$, $L_2 = B_4$, and the orbits X_1, X_2 are Cayley projective planes.*

Proof. From (7.2), we know that there are only three possibilities: (i) $L = N_1 \times N_2$, (ii) $p = q$, N_1 and N_2 are finite, and (iii) $p = q = 4$, N_1 and N_2 are both C_1, and L is locally isomorphic with $C_1 \times C_1 \times C_1$. Let us consider them separately.

(i) In this possibility, we know from (7.3) that either $X_1 = S^q$, $X_2 = S^p$, $X = S^q \times S^p$, or G is of rank 2 and L is the two-dimensional toral group T^2. The first alternative is the case (I) in the above list, and therefore let us assume that G is of rank 2 and $L = T^2$. Since G acts effectively on X, and X has positive Euler characteristic, G has a trivial center, and in particular, is semi-simple. If G is simple, then it is the case (III). If G is not simple, then $G = D_2$, $X = D_2/T^2 = S^2 \times S^2$, whence $X_1 = X_2 = S^2$. This is included in the case (I) of the above list.

(ii) In this case, $p = q$, N_1 and N_2 are finite. Since N_1 is finite, L_1 acts almost effectively on the p-sphere L_1/L, and hence L_1 is either G_2 or a

B_m, where $m = p/2$. We know that $L_1 \subset G$ and G, L_1 have the same rank. If $L_1 = G_2$, then $G = L_1$ (because no other group of rank 2 can contain G_2 as a proper subgroup), whence X_1 is a single point, and we get case (II). Suppose $L_1 = B_m$. We know [3] that there is no compact connecter Lie group of rank m which can contain B_m as a proper subgroup unless $m = 4$, and when $m = 4$, F_4 is the only group having these properties. Therefore, either $G = L_1$, or $(G, L_1) = (F_4, B_4)$. In the first alternative, X_1 is reduced to a point, and we get case (II). In the second alternative, $G = F_4$, $L = B_4$, whence $p = q = 8$, and $L = D_4$, $L_2 = B_4$ on account of (7.1). The orbits X_1, X_2, being homeomorphic with F_4/B_4, are the Cayley projective planes. This is the case (V) in our list.

(iii) In this case, $p = q = 4$, $N_1 = C_1$, $N_2 = C_1$ and L is locally isomorphic with $C_1 \times C_1 \times C_1$. Since L_1/N_1 acts transitively and effectively on the 4-sphere L_1/L, L_1 is locally isomorphic with $C_1 \times C_2$. The only connected and compact Lie group which can contain L_1 as a proper subgroup of the same rank is C_3 [3]. Therefore, either $G = L_1$ or $G = C_3$. In the first alternative, X_1 is a point, and we get case (II). In the second alternative, $G = C_3$, L_1 and L_2 are locally isomorphic with $C_1 \times C_2$. The exceptional orbits X_1 and X_2, being homeomorphic with $C_3/C_1 \times C_2$, are quaternion projective planes. This is the case (IV) in the above list.

(7.5) *In the Case* III *of* (7.4), *the connected center of at least one of* L_1, L_2 *is not contained in the semi-simple part of the other.*

Proof. Let us write $L_1 = N_1M$, $L_2 = M'N_2$, where M, M' are, respectively, the largest connected normal subgroups of L_1, L_2 which act almost effectively on L_1/L, L_2/L. From the proof of (7.4), we know that $L = N_1 \times N_2 = T^2$. Since both L_1 and L_2 are locally isomorphic with $T \times A_1$, it follows that N_1 and N_2 (M and M') are, respectively, the connected center (semi-simple part) of L_1 and L_2. Suppose (7.5) to be false. Then $N_1 \subset M'$, $N_2 \subset M$. By using exactly the same reasoning as in the proof of Case 2 of (7.3), we know that X_1 and X_2 must be spheres. A contradiction is thus obtained, and (7.5) is proved.

8. **The case by which q is even and $p > 1$ is odd.** Suppose that p is odd and q is even. Then L_1/L is an odd sphere and L_2/L an even sphere. Theerfore, in the diagram (5.1), α_1 is surjective and α_2 injective and

$$\begin{aligned} f_2^*(K_2) = f_2^*\sigma_2(I(L_2)) = \sigma\alpha_2(I(L_2)) \subset \sigma(I(L)) \\ = f_1^*\sigma_1(I(L_1)) = f_1^*(K_1), \end{aligned}$$

where K_1, K_2 denote the characteristic subalgebras of the homogeneous spaces $X_1 = G/L_1$, $X_2 = G/L_2$ respectively. From (3.4), the common elements of the images of f_1^* and f_2^* are of zero degree. Therefore, K_2 as well as $f_2^*(K_2)$ contains only elements of zero degree. It follows that h_2^* is injective and L_2 is nonhomologous to zero in G.

Let $\mathfrak{P}(X)$, $\mathfrak{P}(X_1)$, $\mathfrak{P}(X_2)$ be the Poincaré polynomials of X, X_1, X_2 respectively with t as the indeterminate. Then from (3.4) and (3.5), we have

$$\mathfrak{P}(X) = (1+t^p)\mathfrak{P}(X_1) = (1+t^q)\mathfrak{P}(X_2), \tag{8.1}$$

$$\mathfrak{P}(X) = \mathfrak{P}(X_1) + \mathfrak{P}(X_2) + t^{n-1} - 1. \tag{8.2}$$

Let s be the largest integer such that $(1+t)^s$ divides $\mathfrak{P}(X_2)$. Since p is odd, $s \geqq 1$. From (8.1), $(1+t)^{s-1}$ divides $\mathfrak{P}(X)$, $\mathfrak{P}(X_1)$ and $\mathfrak{P}(X_2)$, and so $(1+t)^{s-1}$ divides $t^{n-1}-1$ on account of (8.2). It follows that s is equal to 1 or 2.

Now suppose that $s = 1$. Since $L_2 \not\sim 0$ in G, by Samelson's results, X_2 is a real homology sphere of odd dimension. We have then from (8.1) and (8.2) that $\mathfrak{P}(X_1) = 1 + t^q$, $\mathfrak{P}(X_2) = 1 + t^p$, $p + q = n - 1$. Proposition (4.8) then tells us that $X_1 = S^q$, $X_2 = S^p$, $X = S^p \times S^q$.

Suppose that $s = 2$. Then $\mathfrak{P}(X_2)$ takes the form $(1+t^u)(1+t^v)$, where u, v are odd. Taking account of (8.1) and (8.2), we can get by a direct computation that

$$\mathfrak{P}(X_1) = (1+t^q)(1+t^{p+q}),\ \mathfrak{P}(X_2) = (1+t^p)(1+t^{p+q}),\ 2(p+q) = n-1,$$
$$\mathfrak{P}(X) = (1+t^p)(1+t^q)(1+t^{p+q}).$$

Thus we have proved

(8.3) *Suppose that q is even, p is odd and greater than* 1. *Then either* (A) $X_1 = S^q$, $X_2 = S^p$, $X = S^p \times S^q$, *or* (B) L_2 *is non-homologous to zero in* G, $n = 2(p+q)+1$, *and*

$$\mathfrak{P}(X_1) = (1+t^q)(1+t^{p+q}), \qquad \mathfrak{P}(X_2) = (1+t^p)(1+t^{p+q}).$$

To study Case B in (8.3), we need to analyse the structures of G, L, L_1, L_2 in detail, and for this purpose, we find it convenient to lay down the following

DEFINITION. *Let W be a connected Lie group and $W_1, W_2, \cdots, W_r$ connected subgroups. We say that W is locally the direct product of $W_1, W_2, \cdots, W_r$ if the Lie algebra of W is the direct sum of the Lie algebras*

of the W_i*'s.* If this is the case, we shall write $W = W_1 \cdot W_2 \cdot \cdots \cdot W_r$. We note that each W_i is a normal subgroup of W.

Now let us consider the subgroups L, L_1, L_2. Since both p and q are greater than one, they are connected. Let N_1 be the maximum connected normal subgroup of L_1 which acts trivially on L_1/L. Then we can write $L_1 = M \cdot N_1$, where M acts almost effectively on L_1/L. But L_1/L is an odd sphere, and so by Montgomery-Samelson theorem, we have $M = M_s \cdot J$, where M_s is simple and transitive on L_1/L while J is either the identity or a group of rank 1. Let $M' = M \cap L$, $M_s' = M_s \cap L$. Then $M' = M_s' \cdot J'$, $L = M' \cdot N_1$, where J' is a normal subgroup isomorphic with J. It is is to be noted that if $M_s' = e$, then $J = J' = e$ and L_1/L is a three-sphere. Similarly, we can write $L_2 = Q \cdot N_2$, $L = Q' \cdot N_2$, where Q acts almost effectively and transitively on the even sphere L_2/L while N_2 acts trivially, and $Q' = Q \cap L$.

(8.4) *The subgroups* L, L_1, L_2 *have local direct decomposition as follows:*

$$L_2 = Q \cdot N_2, \quad L_1 = M_s \cdot J \cdot N_1, \quad L = Q' \cdot N_2 = M_s' \cdot J' \cdot N_1,$$

where $J' \subset M_s \cdot J$, $Q/Q' = S^q$, $M_s/M_s' = S^p$ *and* $J \approx J'$ *is of rank* $\leqq 1$. *If, moreover,* $p + q > 15$, *then* M_s' *is a normal subgroup of* N_2 *and* Q' *is a normal subgroup of* $J' \cdot N_1$.

Proof. The first part is merely a summary of what we have just obtained. It remains only to show the second part. If M_s' is the identity, then this is evident. Otherwise, M_s' is a connected simple non-abelian normal subgroup of L. Therefore, M_s' is either a normal subgroup of Q' or that of N_2.[6] But the compact transitive transformation groups of spheres as well as their isotropic subgroups are completely known [1]. A direct verification tells us that when $p + q > 15$, M_s' and Q' cannot have any common normal subgroup of positive dimension. It follows that $M_s' \subset N_2$ and $Q' \subset J' \cdot N_1$.

(8.5) *Suppose that* G *acts almost effectively on* S^n *with an* $(n-1)$-*dimensional orbit, and* G *is a local direct product* $G = G' \cdot G''$, *where* $G' \neq e$, $G'' \neq e$. *If* $L_i = L_i' L_i''$, $L = L'L''$, *where* L', $L_i' \subset G'$ *and* $L'', L_i'' \subset G''$ $(i = 1, 2)$, *then* $X_1 = S^q$, $X_2 = S^p$.

Proof. The direct product $G' \times G''$ is a finite covering of G, and thus acts on S^n almost effectively with the same orbits as G. Moreover, the isotropic subgroups of $G' \times G''$ satisfy the conditions of our proposition (8.5). Therefore, we can simply take G to be the direct product $G' \times G''$. It is evident that

[6] This follows from the fact that $Q' = O(q)$ and $q > 2$.

$L' \subset L_i'$, $L'' \subset L_i''$, $X = (G'/L') \times (G''/L'')$, $X_i = (G'/L_i') \times (G''/L_i'')$, $(i = 1, 2)$. Let

$$j_i' \colon H^*(G'/L_i') \to H^*(G'/L'), \qquad j_i'' \colon H^*(G''/L_i'') \to H^*(G''/L_i'')$$

be the homomorphisms of the real cohomology induced by the inclusions of the isotropic subgroups. We have the following commutative diagram:

$$\begin{array}{ccc}
H^*(X_1) & = & H^*(G'/L_1') \otimes H^*(G''/L_1'') \\
\downarrow{\scriptstyle f_1^*} & & \downarrow{\scriptstyle j_1' \otimes j_1''} \\
H^*(X) & = & H^*(G'/L') \otimes H^*(G''/L'') \\
\uparrow{\scriptstyle f_2^*} & & \uparrow{\scriptstyle j_2' \otimes j_2''} \\
H^*(X_2) & = & H^*(G'/L_2') \otimes H^*(G''/L_2'').
\end{array}$$

Since f_1^*, f_2^* are injective, j_1', j_2', j_1'', j_2'' are all injective. The dimension of X being greater than the dimension of X_1, we can therefore, up to a change of notation, assume that $m = \dim G'/L' > \dim G'/L_1'$. Since X is orientable, there exists a non-zero element $u \in H^m(G'/L')$ of degree m. By our assumptions, $G'' \neq e$ and G acts almost effectively on S^n. Therefore G''/L'' cannot be a point and then $m < \dim X = n - 1$. It follows then from (3.4) that there exists $u_2 \in H^m(G'/L_2')$ such that $j_2'(u_2) = u$, and hence $\dim G'/L_2' \geqq m = \dim G'/L'$. This tells us that $\dim L_2' \leqq \dim L'$, whence $L_2' = L'$ on account of their connectedness. Thus

$$H^*(G'/L') \otimes 1 \subset f_2^*(H^*(X_2)) = j_2'(H^*(G'/L_2')) \otimes j_2''(H^*(G''/L_2''))\,.$$

It follows then from (3.4) that $H^*(G'/L_1') \approx j_1'(H^*(G'/L_1'))$ can contain only elements of zero degree. But X_1 as well as G'/L_1' is orientable, and so $G' = L_1'$. From the equality $L_2' = L'$, we have $\dim G''/L_2'' \leqq \dim G''/L''$, and then we can use the same argument to show that $G'' = L_2''$, $L_1'' = L''$. It follows then

$$\begin{aligned}
\dim X_1 + \dim X_2 &= \dim G''/L_1'' + \dim G'/L_2' \\
&= \dim G''/L'' + \dim G'/L' = \dim X.
\end{aligned}$$

Proposition (4.8) then tells us that $X_1 = S^p$, $X_2 = S^p$ and $X = S^p \times S^q$.

(8.6) *Suppose that $n > 31$, and the Case B in (8.3) happens. If G is almost effective in S^n, then it is also almost effective on X_2.*

Proof. Let us write $G = G' \cdot G_0$ as local direct product of G' and G_0, where G_0 is the maximal connected normal subgroup of G acting trivially

on X_2 and G' a connected normal subgroup of G acting transitively and almost effectively on X_2. Evidently $G_0 \subset L_2$. Suppose that $Q, Q', N_2, M_s, M_s', \cdots$ have the same meaning as in (8.4). Since $L_2 = Q \cdot N_2$, the intersection $G_0 \cap N_2$ is normal in G_0, and therefore normal in G. By assumption, G acts almost effectively on S^n and so almost effectively on $X = G/L$. It follows that $G_0 \cap N_2$ must be finite, whence $G_0 \subset Q$. But Q is simple, and so either $G_0 = e$ or $G = Q$.

We shall show that $G_0 \neq Q$ by the method of contradiction. Suppose that $G_0 = Q$. Let us consider the local direct decompositions:

$$L_1 = M_s \cdot J \cdot N_1, \qquad L = M_s' \cdot J' \cdot N_1 = Q' \cdot N_2, \qquad L_2 = Q \cdot N_2.$$

Since $N_2 \cap G_0$ is finite and N_2 centralizes G_0, we can take our G' such that $N_2 \subset G'$. By assumption, $p + q = (n - 1)/2 > 15$ and so we have, from (8.4), that $M_s' \subset N_2 \subset G'$. Here two possibilities arise. Let us discuss them separately.

(i) $M_s' \neq e$. The group M_s', being connected, must have positive dimension. Therefore the intersection $M_s \cap G'$ is a normal subgroup of M_s with positive dimension, and then from the simplicity of M_s, we have $M_s = M_s \cap G'$, i.e., $M_s \subset G'$. By our choice, M_s is transitive on L_1/L and therefore $L_1 = M_s L = (M_s N_2) \cdot Q$. Moreover, $L_2 = N_2 \cdot Q$, $L = N_2 \cdot Q'$ and $M_s N_2$, $N_2 \subset G'$ and $Q, Q' \subset G_0$. (8.5) then tells us that X_1 and X_2 are spheres and a contradiction is thus obtained.

(ii) $M_s' = e$. It follows that $p = 3$, $J = J' = e$ and

$$L_1 = M_s \cdot N_1, \quad L_2 = Q \cdot N_2, \quad L = N_1 = Q' \cdot N_2, \quad q > 12.$$

The subgroup M_s centralizes N_1 and hence centralizes Q'. But $Q/Q' = S^q$ and $q \neq 2$, and so the centralizer of Q' in Q is finite. It follows then from the connectedness of M_s and the local direct decomposition $G = G' \cdot G_0 = G' \cdot Q$ that $M_s \subset G'$. We have then

$$L_1 = (M_s N_2) \cdot Q', \quad L = N_2 \cdot Q', \quad L_2 = N_2 \cdot Q,$$

where $M_s N_2, N_2 \subset G'$ and $Q, Q' \subset G_0$. (8.5) then tells us that X_1, X_2 are spheres. This contradicts our assumption. Hence $G_0 = e$ and (8.6) is proved.

(8.7) *Suppose that $n > 31$, G acts almost effectively on S^n, and G' is a connected normal semi-simple subgroup of G transitive on X_2. If Case B in* (8.3) *happens, then G has a local direct decomposition $G = G' \cdot W$, where W is at most of rank* 2 *and is locally isomorphic with a subgroup of the centralizer of $L_2 \cap G'$ in G'. Moreover, $L_2 \cap W$ is finite.*

Proof. Since G is compact and connected, we can write $G = G' \cdot W$, where W is a connected normal subgroup of G. Up to a finite covering which does not affect any of our situation, we can assume, moreover, that $G' = G' \times W$. Let $\xi\colon G \to G'$, $\eta\colon G \to W$ be the projections, and $L_2' = L_2 \cap G'$. The connected group L_2 can be written in the form $L_2 = L_2' \cdot U$, where U is a certain connected and normal subgroup of L_2. Since G' is transitive on X_2 and $U \cap G'$ is finite, $W = \eta(U)$ and W is locally isomorphic with U.

The intersection $L_2 \cap W$ is a normal subgroup of L_2. Therefore, $L_2 \cap W$ is a normal subgroup of $\eta(L_2)$ which coincides with W. This tells us that $L_2 \cap W$ acts trivially on X_2. By (8.6), G is almost effective on X_2 and so $L_2 \cap W$ must be finite. This implies that $\xi(U)$ is locally isomorphic with U and hence with W. The subgroup U centralizes L_2' and therefore, $\xi(U)$ centralizes L_2'. From (8.3), we know that $L_2' \not\sim 0$ in G' and $r(L_2') = r(G') - 2$, where "r" denotes the rank. Moreover L_2' is semi-simple because G' is so. Hence the centralizer of L_2' is of rank 2 at most. Thus $r(W) = r(\xi(U)) \leqq 2$.

9. Some properties of G. In this section, we shall establish some properties of G which lead to the determination of the system $\{G, L, L_1, L_2\}$ when Case B in (8.3) happens. For this purpose, let us digress a little and prove two simple propositions.

(9.1) *Let W be a closed connected subgroup of a connected and compact Lie group V such that $r(W) = r(V) - s$, r denoting the rank. For any normal subgroup V' of V, $0 \leqq r(V') - r(V' \cap W) \leqq s$. Suppose W' to be a closed connected, normal non-abelian simple subgroup of W with $r(W') > s$. Then W' is contained in a certain simple normal subgroup of V.*

Proof. Consider the projection $\eta\colon V \to V/V'$. We have evidently

$$r(V/V') = r(V) - r(V'), r(\eta(W)) = r(W) - r(V' \cap W),$$
$$r(\eta(W)) \leqq r(V/V').$$

It follows that

$$\begin{aligned} r(V') - r(V' \cap W) &= r(V) - r(V/V') - r(W) + r(\eta(W)) \\ &\leqq r(V) - r(V/V') - r(W) + r(V/V') \\ &= r(V) - r(W) = s. \end{aligned}$$

To show the second part of (9.1), let us write $V = V_1 \cdot V_2 \cdot \cdots \cdot V_a$ as local direct product of simple normal subgroups, and

$$\xi_i : V \to V/V_1V_2 \cdots V_{i-1}V_{i+1} \cdots V_a$$

be the projection $(i = 1, 2, \cdots, a)$. There exists at least one i, say 1, such that $\xi_i(W') \neq e$. Since W' is simple, $\xi_1(W')$ is locally isomorphic with W', and hence $r(\xi_1(W')) > s$. Let U be the connected component of the intersection $V_1 \cap W$. Since both U and W' are connected normal subgroups of W and W' is simple, either $W' \subset U$ or W' centralizes U. In the first alternative, $W' \subset V_1$ and our proposition is proved. Now let us show that the second alternative cannot happen. Suppose that W' centralizes U. Then $\xi_1(W')$ centralizes $\xi_1(U)$ in $\xi_1(V_1)$. But W' is simple and non-abelian and $r(\xi_1(W')) > s$, so the centralizer of $\xi_1(W')$ in $\xi_1(V_1)$ is of rank smaller than $r(\xi_1(V_1)) - s$. We have therefore $r(\xi_1(U)) < r(\xi(V_1)) - s$. Restricted to V_1, the homomorphism ξ_1 is a local isomorphism. It follows then $r(U) < r(V_1) - s$. On the other hand, we have from the first part of our Proposition (9.1) that $r(V_1) - r(U) = r(V_1) - r(V_1 \cap W) \leqq s$. A contradiction is thus obtained, and the proof is completed.

(9.2) *Let $V = V_1 \times V_2 \times \cdots \times V_a$ be a direct product of connected, compact simple Lie groups, and Y a connected normal subgroup of V with $r(V) = r(Y) + s$. If $r(V_i) > s$ for all i and Y is non-homologous to zero in V, then*

$$Y = (V_1 \cap Y) \times (V_2 \cap Y) \times \cdots \times (V_a \cap Y).$$

Proof. Let $Y_i = V_i \cap Y$. From the first part of (9.1), it follows that $r(Y_i) \geqq r(V_i) - s \geqq 1$. Since $Y \not\sim 0$ in V, the number of simple local direct factors is less than a, and therefore the normal subgroups $Y_1, Y_2, \cdots, Y_a$ of Y must exhaust all its direct factors. Hence $Y = Y_1 \times Y_2 \times \cdots \times Y_a$.

(9.3) *Suppose that $n > 31$ and Case B in (8.3) happens. Then G has a semi-simple connected normal subgroup G' such that G' is transitive on X_2 and that G' has at most two simple director factors.*

Proof. From (8.3), we know that $h_2^* : H^*(G/L_2) \to H^*(G)$ is injective and its image is generated by two primitive elements x, y of degrees p and $p+q$ respectively. Let us write G as the local direct product $G_1 \cdot G_2 \cdots G_a$ of connected simple normal subgroups. There exists at least a simple factor, say G_1, such that, under the homomorphism $j_1^* : H^*(G) \to H^*(G_1)$ induced by the injection $j_1 : G_1 \to G$, the image $x_1 = j_1^*(x)$ of x is different from zero. Here two alternatives arise. Let us discuss them separately.

(i) Suppose that $y_1 = j_1^*(y) \neq 0$. Consider the commutative diagram

$$\begin{array}{ccc} G/L_2 & \xleftarrow{h_2} & G \\ {\scriptstyle l}\uparrow & & \uparrow{\scriptstyle j_1} \\ G_1/(G_1 \cap L_2) & \xleftarrow{k} & G_1 \end{array}$$

and its induced diagram on the real cohomology algebras

$$\begin{array}{ccc} H^*(G/L_2) & \xrightarrow{h_2^*} & H^*(G) \\ {\scriptstyle l^*}\downarrow & & \downarrow{\scriptstyle j_1^*} \\ H^*(G_1/(G_1 \cap L_2)) & \xrightarrow{k^*} & H^*(G_1), \end{array}$$

where k is the projection and l is the map induced by the inclusion $G_1 \cap L_2 \subset L_2$. Thus $k^*(H^*(G_1/(G_1 \cap L_2))$ contains two non-zero primitive elements x_1, y_1 of degrees p, $p+q$ respectively. It follows that

$$\dim(G_1/(G_1 \cap L_2)) \geqq 2p + q = \dim G/L_2.$$

On the other hand, $G_1/(G_1 \cap L_2)$ can be regarded as the orbit of G_1 when it acts on G/L_2. Therefore G_1 is transitive on X_2. (9.3) is proved in this case.

(ii) Suppose that $j_1^*(y) = 0$. There exists another simple factor, say G_2, such that $j_2^*(y) \neq 0$, where j_2 is the injection: $G_2 \to G$. Just as before, but replacing G_1 by $G_1 \cdot G_2$, we can show that $G_1 \cdot G_2$ is transitive on X_2. This completes the proof.

(9.4) *Suppose that* $n > 31$ *and Case B in* (8.3) *happens. Then* G *cannot be simple.*

Proof. Suppose G to be simple. Let $Q, Q', N_2, M, M_8', \cdots$ have the same meaning as in (8.4). Then $L_2 = Q \cdot N_2$, where Q acts transitively and almost effectively on the q-sphere L_2/L. Since $L_2 \not\sim 0$ in G and G is simple, L_2 must be simple and hence $N_2 = e$. By (8.4), $M_8' \subset N_2 = e$, and then $p = 3$. But we know that $p + q = (n-1)/2 > 15$. Therefore $q > 12$ and Q must be a B_m with $m = q/2 > 6$. Thus the simple group G has the properties: (a) $B_m \not\sim 0$ in G, (b) $r(G) = m + 2$, and (c)

$$\mathfrak{P}(G) = \mathfrak{P}(B_m)(1 + t^3)(1 + t^{2m+3}), \quad m > 6,$$

where $\mathfrak{P}$ denotes the Poincaré polynomial with t as the indeterminate. By a direct enumeration of simple groups, we find that no such G exists. Our proposition is thus proved.

(9.5) *Suppose that $n > 31$ and Case B in (8.3) happens. If G has no proper normal subgroup transitive on X, then G is the local direct product of two simple Lie groups.*

Proof. Without loss of generality, we can assume that G is the direct product of simple Lie groups. On account of (9.4), it suffices to show that G has at most two simple factors. Let G' be the normal subgroup of G as contructed in (9.3), and let us write $G = G' \times W$. Since Q is transitive on L_2/L and G' transitive on X_2, we have $QL = L_2$, $G'QL = G'L_2 = G$. This tells us that $G'Q$ is transitive on X. Let $\eta: G \to W$ be the projection. Then $\eta(L_2) = W$ and $G'Q = \eta^{-1}(\eta(Q))$. Since Q is normal in L_2, $\eta(Q)$ is normal in W, whence $G'Q$ is normal in G. Therefore $G = G'Q$ on account of our assumption. If $Q \subset G'$, then $G = G'$ and by (9.3), G has at most two simple factors. If G' is simple, then G is locally isomorphic with the direct product $G' \times Q$ which has two simple factors. Thus we have proved our proposition when either $Q \subset G'$ or G' is simple.

In fact, the above two cases are the only alternatives. Let us show this by the method of contradiction. Suppose that $Q \not\subset G'$ and G' is not simple. By (9.3), we can write $G' = V_1 \times V_2$, where V_1 and V_2 are simple and none of them is transitive on X_2. Let $L_2' = L_2 \cap G'$. Then $X_2 = G'/L_2'$. Since X_2 is simply-connected and G' connected, the subgroup L_2' must be connected. The group Q being simple and $Q \not\subset G'$, it follows that $\eta(Q)$ is locally isomorphic with Q. From (8.7), we know that $r(W) \leqq 2$ and then

$$r(Q) = r(\eta(Q)) \leqq r(W) \leqq 2.$$

Let $\xi: G \to G'$, $\xi_1: G \to V_1$, $\xi_2: G \to V_2$ be the projections. From (8.7), $\xi(L_2)$ is locally isomorphic with L_2. Now we find it convenient to divide our further discussions into two caeses.

Case I. $r(Q) = 2$. Then $r(W) = 2$ and

$$r(G') = r(G) - r(W) = r(G) - 2 = r(L_2) = r(\xi(L_2)).$$

The subgroup $\xi(L_2)$, being of the same rank as G', can be written as the direct product of a subgroup in V_1 and a subgroup in V_2 [14, p. 927]. Since $L_2 \not\sim 0$ in G, $L_2' \not\sim 0$ in G'. Hence L_2' is a connected, normal semi-simple subgroup of L_2 and hence we can write $L_2' = Y_1 \times Y_2$, where $Y_1 \subset V_1$, $Y_2 \subset V_2$. Evidently, Y_1, Y_2 are connected. Moreover, from the facts $r(L_2') = r(G') - 2$ and $L_2' \not\sim 0$ in G', it follows that

$$Y_1 \not\sim 0 \text{ in } V_1, \quad Y_2 \not\sim 0 \text{ in } V_2, \; r(Y_1) + r(Y_2) = r(V_1) + r(V_2) - 2.$$

By (8.7), $\xi(Q) \neq e$, and therefore either $\xi_1(Q)$ or $\xi_2(Q)$ is non-trivial, say $\xi_1(Q) \neq e$. Q and Y_1 are two connected normal subgroups of L_2 with $Q \cap Y_1$ finite. It follows that Q and Y_1, and hence $\xi_1(Q)$ and Y_1 centralize each other. But $\xi_1(Q)$ is simple and of rank 2, so

$$r(V_1) \geqq r(Y_1 \cdot \xi_1(Q)) = r(Y_1) + r(\xi_1(Q)) = r(Y_1) + 2.$$

We have therefore $r(Y_2) = V_2$ and then $Y_2 = V_2$. This contradicts the fact that V_1 is not transitive on X_2. In other words, this case cannot happen.

Case II. $r(Q) = 1$. From the construction of V_1; V_2 in (9.3), we know that one of V_1, V_2 has a primitive exponent[7] p while the other a primitive exponent $p+q$. Since $r(Q) = 1$ and $2(p+q) = n > 31$, we have $q = 2$, $p > 13$, whence

$$r(V_1) \geqq 4, \qquad r(V_2) \geqq 4.$$

Let $Y_1 = L_2' \cap V_1 = L_2 \cap V_1$, $Y_2 = L_2' \cap V_2 = L_2 \cap V_2$. Since $L_2' \nsim 0$ in G' and $r(L_2') = r(G') - 2$, it follows from (9.2) that $L_2' = Y_1 \times Y_2$, whence $X_2 = G'/L_2' = (V_1/Y_1) \times (V_2/Y_2)$. We know that X_2 is simply connected, neither V_1 nor V_2 is transitive on X_2 and $H^*(X_2) \approx H^*(S^p \times S^{p+q})$. It follows that V_1/Y_1 and V_2/Y_2 are simply connected homology spheres with real coefficients. One of them has dimension p while the other dimension $p+2$. The integers p, $p+2$ cannot be both congruent to $3 \bmod 4$. To be precise, let us take $\dim V_1/Y_1 \not\equiv 3 \pmod 4$. Y_1 and Q are two connected normal subgroups of L_2 with $Q \cap Y_1$ finite. It follows that Y_1 and Q, and hence Y_1 and $\xi_1(Q)$, centralize each other. But $\xi_1(Q)$ is simple and so $Y_1 \cap \xi_1(Q)$ is a finite central subgroup of $\xi_1(Q)$. Now let use consider the fibring $V_1/Y_1 \to V_1/(Y_1 \cdot \xi_1(Q))$ whose fibres are homeomorphic with $F = \xi_1(Q)/Y_1 \cap \xi_1(Q)$. Since $r(Q) = 1$, F is either a point, a 3-sphere or a 3-dimensional real projective space. We know that V_1/Y_1 is a simply-connected real homology sphere of dimension $\not\equiv 3 \pmod 4$. It cannot have any orientable fibre decomposition with the 3-dimensional real homology sphere as the fibre. Hence F must be a point and $\xi_1(Q) = e$, or in other words, $Q \subset V_2 \times W$. Let us denote $V_2 \times W$ by W', and let $Q', N_2, M_s, \cdots$ have the same meaning as in (8.4). Since $G = G'Q$, $L_2 = L_2' \cdot Q$ and $L = L_2' \cdot Q'$, we have

$$G = V_1 \times W',\ L_2 = Y_1 \times Y_2 Q,\ L = Y_1 \times Y_2 Q',\ Y_1 \subset V_1 \tag{9.6}$$
$$Y_2 Q' \subset Y_2 Q \subset W'.$$

[7] The degree of a primitive element of the real cohomology of a compact Lie group is called a primitive exponent of that group.

Since $p > 13$, $r(M_s) \geqq 4$, $r(L_1) = r(G) - 1$, it follows from (9.1) that M_s belongs to a simple factor of G. Therefore, either $M_s \subset V_1$ or $M_s \subset W'$, and then

$$L_1 = M_s L \subset (L_1 \cap V_1) \times (L_1 \cap W') \subset L_1. \tag{9.7}$$

Taking account of (9.6) and (9.7), we know from (8.5) that X_1, X_2 are spheres. A contradiction is thus obtained and (9.5) is proved.

(9.8) *Suppose that $n > 31$ and Case B in (8.3) happens. Then $q \leqq 4$ and the group G has a simple normal subgroup G' transition on X_2.*

Proof. Let G^* be a minimum normal subgroup of G transitive on X. Then G and G^* have the same orbits on S^n, and every connected normal subgroup of G^* is a normal subgroup of G. Therefore, without affecting any of our situation, we can simply take G to be G^*. Furthermore, without loss of generality, we can assume that G is a direct product of simple Lie groups. By (9.5), $G = V_1 \times V_2$, where V_1, V_2 are simple. Since X is simply-connected, the semi-simple part of G is transitive on X [15, p. 4], and so G must be semi-simple and V_1, V_2 non-abelian. Now let us divide our discussions into two cases.

Case I. $p = 3$. Then $q > 12$, $Q = B_m$, where $m = q/2 > 6$. The subgroup $L_2 = N_2 \cdot Q$ is non-homologous to zero in $V_1 \times V_2$, and

$$\mathfrak{P}(V_1)\mathfrak{P}(V_2) = \mathfrak{P}(Q)\mathfrak{P}(N_2)(1 + t^3)(1 + t^{3+q}),$$

where $\mathfrak{P}$ denotes the Poincaré polynomial. It follows that N_2 has vanishing first and third Betti number, and hence $N_2 = e$. (8.4) then tells us that $M_s' = e$, and then

$$J \approx J' = e,\ M_s = B_1,\ L_2 = Q,\ L_1 = M_s \cdot Q',\ L = Q'.$$

Taking account of the facts $r(Q) = r(L_2) = r(G) - 2$ and $r(Q) > 2$, we have from (9.1) that Q belongs to a simple normal subgroup of G, say $Q \subset V_2$. Therefore, $X_2 = V_1 \times (V_2/Q)$. By (8.6), G is almost effective on X_2 whence $V_2 \neq Q$. Since V_1 is simple and non-trivial, it follows from the equality $\mathfrak{P}(X_2) = (1 + t^3)(1 + t^{3+q})$ that $\mathfrak{P}(V_2/Q) = 1 + t^{3+q}$. Hence

$$\mathfrak{P}(V_2) = \mathfrak{P}(Q)(1 + t^{3+q}) = \mathfrak{P}(B_m)(1 + t^{3+q}), \quad m = q/2 > 6.$$

An enumeration of simple groups tells us that no such simple V_1 exists. In other words, the case by which $p = 3$ cannot happen.

Case II. $p > 3$. Since $L_2 = Q \cdot N_2$ is non-homologous to zero in $V_1 \times V_2$,

N_2 must be simple. By (8.4), M_s' is a normal subgroup of N_2. Here $p > 3$, $M_s' \neq e$ and so $M_s' = N_2$, whence

$$(9.9) \qquad L_2 = Q \cdot M_s', \ L_1 = M_s \cdot J \cdot N_1, \ L = Q' \cdot M_s', \ Q' = J' \cdot N_1.$$

Case IIa. Suppose that $q > 4$. Then Q' is a simple group of rank greater than or equal to two. Since $r(J') \leqq 1$, the last equality in (9.9) tells us that $J' = e$, and then $J = e$, $Q' = N_1$ and

$$L_2 = Q \cdot M_s', \ L_1 = Q' \cdot M_s, \ L = Q' \cdot M_s'.$$

From the facts $p > 3$, $q > 4$, we have $r(M_s) > 1$, $r(Q') > 1$. Moreover, $r(L_1) = r(G) - 1$ by (8.3). It follows then from (9.1) that both M_s and Q' belong to certain simple normal subgroups of G. Suppose that M_s, Q' belong to a same simple normal subgroup, say V_1. Since Q is simple and $Q' \subset Q$, Q also belongs to V_1, and then $L_2 = Q \cdot M_s' \subset V_1$. The last inclusion would imply that $L_2 \not\sim 0$ in V_1 which is impossible as L_2 has two simple normal factors. Therefore, M_s and Q' belong to different simple normal subgroups of G, say $M_s \subset V_1$, $Q' \subset V_2$. It follows that $Q \subset V_2$, and then (8.5) tells us that X_1 and X_2 are spheres. This contradicts our hypothesis. In other words, q cannot be greater than 4.

Case IIb. Suppose that $q \leqq 4$. We shall prove the second part of our proposition by method of contradiction. Suppose that G has no simple normal subgroup transitive on X_2. From the proof of (9.3), we know that one of V_1, V_2 has a primitive exponent p while the other has a primitive exponent $p + q$. Since $q \leqq 4$, $n > 31$, we have $p > 11$ and

$$r(V_1) \geqq 3, \qquad r(V_2) \geqq 3.$$

Taking account of the fact $r(L_2) = r(G) - 2$, we know from (9.1) that

$$r(V_1 \cap L_2) \geqq 1, \qquad r(V_2 \cap L_2) \geqq 1.$$

But $V_1 \cap L_2$, $V_2 \cap L_2$ are normal subgroups of L_2, and L_2 has only two proper normal subgroups. Therefore, up to a change of notation, we can assume that $M_s' \subset V_1$, $Q \subset V_2$. Since $M_s' \neq e$ and M_s is simple, it follows that $M_s \subset V_1$ and then

$$L_1 = M_s L = M_s \cdot Q', \ L_2 = M_s' \cdot Q, \ L = M_s' \cdot Q',$$

where $Q' \subset Q \subset V_2$, $M_s' \subset M_s \subset V_1$. (8.5) then tells us that X_1, X_2 are spheres. A contradiction is thus obtained. In other words, G has a simple normal subgroup transitive on X_2.

10. Determination of G when p is odd and q is even. Now we are in a position to determine the group G and the orbits X, X_1, X_2 for even q and odd $p > 1$. Let $Sp(r)$ denote the group of all $(r \times r)$ quaternion symplectic matrices. If a subgroup V of $Sp(r)$ is conjugate to the subgroup of all matrices $q = (q_{ij}) \in Sp(r)$ with $q_{11} = 1$, then we shall call V an $Sp(r-1)$ imbedded in $Sp(r)$ in the canonical position. Canonical imbedding for unitary groups, special unitary groups and the orthogonal groups is defined similarly.

(10.1) *Suppose that $n > 31$ and Case B in (8.3) happens. If $q = 4$, then X_2 is the quaternion Stiefel manifold $Sp(r)/Sp(r-2)$, $X = X_2 \times S^4$, and X_1 is an S^{p+4}-bundle over S^4, where $p = 4r - 5$. If the group G is almost effective on S^n, then up to a finite covering of G*

$$G = Sp(r) \times Sp(2),\ Q \approx Sp(2),\ Q \cap Sp(r) = Q \cap Sp(2) = e,$$
$$M_s \approx Sp(r-1),\ M_s' \approx Sp(r-2),$$

where the imbeddings $M_s' \subset M_s$, and $M_s \subset Sp(r)$ are in the canonical manner.

Proof. Up to a finite covering, we can assume that G is a direct product of simply-connected compact simple Lie groups and circle groups. Now let $L, L_1, L_2, Q, Q', M_s, M_s', \cdots$ have the same meaning as before. Since $n > 31$, $q = 4$ and Case B in (8.3) happens, we know that $p > 11$ and $Q = C_2$. From (9.8), G has a simple normal subgroup G' transitive on X_2, and from (8.7), we can write $G = G' \times W$. The subgroup Q cannot be contained in G' for, if so, G' would be transitive on X and would have the same orbits as G which contradicts (9.4). Therefore, $Q \cap G'$ is finite. Let $\eta\colon G \to W$ be the projection. Then $\eta(Q)$ is locally isomorphic with Q and hence a C_2. By (8.7), W is a compact Lie group of rank not greater than 2. We know that there is no connected group of rank 2 which can contain C_2 as a proper subgroup, and so $\eta(Q) = W$. But the fundamental group of G has no torsion. It follows then W is simply-connected, and then $W = Sp(2)$, $W \approx Q$, $Q \cap G' = e$. Since G is the direct product of two simple non-abelian groups, all the arguments in Case II of (9.8) hold, and then

$$(10.2) \qquad L_2 = M_s' \cdot Q,\ L_1 = M_s \cdot Q',\ L = M_s' \cdot Q',\ Q' = J' \cdot N_1.$$

The group $\eta(M_s')$ centralizes $\eta(Q)$ and $\eta(Q) = W = Sp(2)$. It follows that $\eta(M_s') = e$, or what is the same, $M_s' \subset G'$, whence $M_s \subset G'$. Therefore M_s' is the identity component of $L_2 \cap G'$. But $X_2 = G/L_2 = G'/(L_2 \cap G')$ is simply connected, and so $L_2 \cap G'$ is connected. We have then $M_s' = L_2 \cap G'$ and $X_2 = G'/M_s'$. Let $\xi\colon G \to G'$ be the projection. By (8.7), L_2 is locally isomorphic with $\xi(L_2) = M_s' \cdot \xi(Q)$, and, from (8.3),

$$r(G') = r(G) - r(W) = r(G) - 2 = r(L_2) = r(\xi(L_2)).$$

We know thus that the pair $\{G', M_s'\}$ has the following five properties:

(i) G', M_s' are simple groups,

(ii) M_s' is non-homologous to zero in G',

(iii) $M_s/M_s' = S^p$, $p = \text{odd} > 11$,

(iv) $\mathfrak{P}(G'/M_s') = \mathfrak{P}(X_2) = (1 + t^p)(1 + t^{p+4})$, and

(v) G' has a subgroup $M_s' \cdot \xi(Q)$ having the same rank as G', where $\xi(Q)$ is a C_2.

By a direct enumeration of simple groups, we find that

$$G' = C_r,\ M_s' = C_{r-2},\ M_s = C_{r-1},\ p = 4r - 5.$$

Since $\pi_1(G)$ has no torsion, $G' = Sp(r)$, $G = Sp(r) \times Sp(2)$.

Now let us discuss the position of M_s in G', and the position of M_s' in M_s. We recall that

$$L_1 = M_s \cdot J \cdot N_1,\ L = M_s' \cdot J \cdot N_1,\ Q' = J' \cdot N_1 \approx C_1 \times C_1,$$
$$J' = M_s \cdot J,\ r(J) = r(J') \leqq 1.$$

Since $r(G) = r(L_1) + 1$,

$$r(L_1 \cap W) = r(L_1) - r(\xi(L_1)) \geqq r(L_1) - r(G') = r(L_1) - r(G) + 2 = 1.$$

On the other hand, $M_s \subset G'$, $N_1 \cap W \subset Q \cap W = e$. It follows that J is the identity component of $L_1 \cap W$ and $r(J) = 1$, $N_1 \approx C_1$. Therefore, $r(L_1 \cap W) = 1$, and the subgroup $\xi(L_1) = M_s \cdot \xi(N_1)$ has the same rank as G'. Up to an automorphism of $Sp(r)$, there is only one connected subgroup of the class C_{r-1} which is a normal subgroup of a subgroup of maximum rank in $Sp(r)$ [14]. Hence M_s is an $Sp(r-1)$ imbedded in $G' = Sp(r)$ in the canonical position. We know that M_s/M_s' is a p-sphere. Up to an automorphism of M_s, there is only one such subgroup M_s' [17].[8] Therefore, M_s' is a $Sp(r-2)$ imbedded in $M_s = Sp(r-1)$ in the canonical position.

We are now in a position to determine the orbits X, X_1, X_2. Obviously, $X_2 = G'/M_s' = Sp(r)/Sp(r-2)$ is the quaternion Stiefel manifold. To see X, let us consider the map

$$\theta\colon (G'/M_s') \times (Q/Q') \to G/L = X$$

given by $\theta(gM_s', qQ') = gqL$, where $g \in G'$, $q \in Q$. We can easily verify that

[8] This is equivalent to saying that M_s acts on the p-sphere M_s/M'_s orthogonally.

this mapping θ is well defined. From the fact that $G' \cap Q = e$, θ must be injective. Moreover, $(G'/M_s') \times (Q/Q')$ and X have the same dimension, and so θ is bijective. This tells us that X is homeomorphic with $X_2 \times S^4$. To see X_1, let us consider G' to be a transformation group of $X_1 = G/L_1$. Since G' is normal in G, the G'-orbit decomposition of X_1 is a fibre decomposition. The fibres (i. e., orbits under G') are all homeomorphic with

$$G'/(G' \cap L_1) = G'/M_s = Sp(r)/Sp(r-1) = S^{4r-1} = S^{p+4}$$

whereas the base space (i. e., the space of orbits) is

$$G/(G'L_1) = W(W \cap L_1) = W/\eta(Q') = Q/Q' = S^4.$$

Therefore X_1 is an S^{p+4}-bundle over S^4. This completes the proof of (10.1).

Remark. G does not act effectively on S^n. There is a central subgroup I_2 of order 2 which is contained in neither direct factor of G. The factor group G/I_2 acts effectively on S^n.

(10.3) *Let G, $\bar{G}$ be two compact connected effective differentiable transformation groups of S^n with $(n-1)$-dimensional orbit. Suppose that $n > 31$, p is odd, $q = 4$, and their special orbits are not spheres. Let L, L_1, L_2 and $\bar{L}$, $\bar{L}_1$, $\bar{L}_2$ be the isotropic subgroups of G and $\bar{G}$ respectively defined as above. Then there is an isomorphism of G onto G carrying L, L_1, L_2 into $\bar{L}$, $\bar{L}_1$, $\bar{L}_2$ respectively.*

Proof. The universal covering groups of G, $\bar{G}$ act on S^n in the natural manner, and have the same orbits as G, $\bar{G}$ respectively. We can readily see that if these two covering groups satisfy the conclusion of (10.3), so do G and $\bar{G}$. Hence it is sufficient for us to prove our proposition for the case that G and $\bar{G}$ are simply-connected. Of course, G and $\bar{G}$ are now no longer effective. From (10.1),

$$G = \bar{G} = Sp(r) \times Sp(2).$$

Suppose that M_s, M_s', Q, Q' are subgroups of G defined by (8.4), and $\bar{M}_s$, $\bar{M}_s'$, $\bar{Q}$, $\bar{Q}'$ are subgroups of $\bar{G}$ similarly defined. From (10.1), the pair (M_s, M_s') and $(\bar{M}_s, \bar{M}_s')$ are conjugate in $G' = Sp(r)$, and thus we can assume

$$M_s = \bar{M}_s = Sp(r-1), \qquad M_s = \bar{M}_s' = Sp(r-2),$$

where the imbeddings $Sp(r-2) \subset Sp(r-1)$, $Sp(r-1) \subset Sp(r)$ are in the canonical manner. Let C be the identity component of the centralizer of M_s' in $Sp(r)$. Then $C \approx Sp(2)$. Since $C \times Sp(2)$ is the identity component of the centralizer of M_s' in G, $Q \subset C \times Sp(2)$. From (10.1), $Q \cap C = Q \cap Sp(2)$

$= e$. Therefore, Q is a diagonal of $C \times Sp(2)$. Up to an automorphism of $Sp(2)$, there is only one diagonal of $C \times Sp(2)$. Since $\bar{Q}$ is also a diagonal of $C \times Sp(2)$ and this automorphism of $Sp(2)$ does not affect $Sp(r)$, we can further assume that $Q = \bar{Q}$.

We know from (10.2) that Q' and M_s centralize each other. Since M_s is an $Sp(r-1)$ imbedded in $Sp(r)$ in the canonical position and Q is a diagonal of $C \times Sp(2)$, we can verify readily that Q' is the maximum connected subgroup of Q which centralizes M_s. Therefore, Q' is completely determined by M_s and Q. From the equalities $M_s = \bar{M}_s$, $Q = \bar{Q}$, it follows that $Q' = \bar{Q}'$. (10.2) then tells us that $L = \bar{L}$, $L_1 = \bar{L}_1$, $L_2 = \bar{L}_2$. Proposition (10.3) is thus proved.

(10.4) *Suppose that $n > 31$, $q = 2$, p is odd, and Case B in* (8.3) *happens. Then X_2 is the complex Stiefel manifold $SU(r)/SU(r-2)$, $X = X_2 \times S^2$, and X_1 is an S^{p+2}-bundle over S^2, where $p = 2r-3$. If, moreover, G is semi-simple, simply-connected and almost effective on S^n, then*

$$G = SU(r) \times SU(2),\ Q \approx SU(2),\ Q \cap SU(r) = Q \cap SU(2) = 2,$$

$$M_s = SU(r-1),\ M_s' = SU(r-2),$$

$$L_2 = M_s' \cdot Q,\ L_1 = M_s \cdot Q',\ L = M_s' \cdot Q',$$

and the imbeddings $M_s \subset SU(r)$, $M_s' \subset M_s$ are canonical.

Proof. The proof is similar to that of (10.1).

(10.5) *Suppose that $n > 31$, $q = 2$, p is odd, and Case B in* (8.3) *happens. If G acts almost effectively on S^n, then G is either semi-simple or a local direct product of a circle group and a semi-simple group.*

Proof. Let us write G in the form of local direct product $G^* \cdot W$, where G^* denotes the semi-simple part and W the identity component of the center of G. Since X, X_1, X_2 are all simply-connected, G and G^* have the same orbits on S^n [15, p. 4]. By (8.7), W is locally isomorphic with a subgroup of the centralizer of $L_2 \cap G^*$ in G^*. On the other hand, by applying (10.4) to G^*, we know that the centralizer of $L_2 \cap G^*$ in G^* is one-dimensional. Therefore W is either trivial or the circle group, and (10.5) is thus proved.

(10.6) *Let G, $\bar{G}$ be two compact connected effective differentiable transformation groups of S^n with an $(n-1)$-dimensional orbit. Suppose that $n > 31$, $q = 2$, p is odd, and their special orbits are not spheres. Let L, L_1, L_2 and $\bar{L}$, $\bar{L}_1$, $\bar{L}_2$ be the isotropic subgroups of G and $\bar{G}$ respectively defined as*

above. If G and $\bar{G}$ are locally isomorphic, then there is an isomorphism of G onto $\bar{G}$ carrying L, L_1, L_2 into $\bar{L}$, $\bar{L}_1$, $\bar{L}_2$ respectively.

Proof. The groups G, $\bar{G}$ can be semi-simple or a local direct product of a circle group and a semi-simple group. We first prove the semi-simple case and then use the result to prove the other case. The argument is similar to the argument used in the proof of (10.3), and we omit the details.

Remark. The assumptions are the same as in (10.6). If G is semi-simple, then G is isomorphic with $SU(r) \times SU(2)$ or $(SU(r) \times SU(2))/I_2$ according as $p \equiv 3$ or $1 \pmod 4$, where I_2 denotes a central subgroup of $SU(r) \times SU(2)$ which belongs to neither factor. If G is not semi-simple, then G is isomorphic with the factor group $(U(r) \times U(2))/T$, where $U(k)$ denotes the unitary group in k variables and T denotes a closed one-dimensional central subgroup of $U(r) \times U(2)$ which intersects neither of the factors.

11. The case by which p or q equals 1. When p or q is equal to 1, the isotropic subgroups L, L_1, L_2 may not be connected. Our previous method to study the orbits of G has to be modified. In this section, we shall first compare the cohomology of X, X_1, X_2 with their covering spaces and then determine them.

(11.1) *Let L^0 denote the identity component of L. Then* $\mathrm{Ad}\, L = \mathrm{Ad}\, L^0$, *where "Ad" denotes the linear adjoint representation on its Lie algebra.*

Proof. If $p > 1$ and $q > 1$, then $L^0 = L$, and then this becomes trivial. Now let us assume $p = 1$. Suppose $q > 1$. Since X is an S^q-bundle over X_2, the homomorphism $f_2^{\#}\colon \pi_1(X) \to \pi_1(X_2)$ is bijective. By (4.3), X_1 must be simply-connected, and hence L_1 is connected. But L_1/L is a circle, and so $L_1 = T \cdot L^0$, where T is a one-dimensional central subgroup of L_1. From the fact that $L \subset L_1 = T \cdot L^0$, it follows immediately that $\mathrm{Ad}\, L = \mathrm{Ad}\, L^0$. Thus (11.1) is proved when $q > 1$.

Now suppose that $p = q = 1$. Then L_1/L, L_2/L are circle groups, and therefore $L_i = T_i \cdot L$, where T_i is a circle group belonging to the center of L_i $(i = 1, 2)$. Let L_i^0 denote the identity component of L_i. Since, for $i = 1, 2$, $L_i^0 \cap L \subset L_i^0 = T_i \cdot L^0$, we have

$$\mathrm{Ad}(L_1^0 \cap L) = \mathrm{Ad}\, L^0, \qquad \mathrm{Ad}(L_2^0 \cap L) = \mathrm{ad}\, L^0.$$

Proposition (4.9) then implies that $\mathrm{Ad}\, L = \mathrm{Ad}\, L^0$. The proof is thus completed.

(11.2) *The homomorphism $\theta^*: H^*(G/L) \to H^*(G/L^0)$ induced by the covering map $\theta: G/L^0 \to G/L$ is bijective.*

Proof. Since the coefficient field is the field of real numbers, θ^* is injective and the image of θ^* consists of the elements of $H^*(G/L^0)$ invariant under the group Σ of covering transformations. Let us recall that, for any Lie group Q (not necessarily connected), $I(Q)$ denotes the algebra of symmetric multilinear forms over the Lie algebra of Q invariant under $\mathrm{Ad}\, Q$. From Cartan's results [4], we know that there is a canonical isomorphism between $H^*(G/L^0)$ and $H^*(I(L^0) \otimes H^*(G))$, where the differential operator over $I(L^0) \otimes H^*(G)$ depends only on L^0, G and the position of L^0 in G. Let us identify $H^*(G/L^0)$ with $H^*(I(L^0) \otimes H^*(G))$. Since the group Σ of covering transformations is finite, we see immediately that $H^*(I(L) \otimes H^*(G))$ consists of all the Σ-invariant elements in $H^*(I(L^0) \otimes H^*(G))$. Therefore, $\theta^*(H^*(G/L)) = H^*(I(L) \otimes H^*(G))$. But from (11.1), it follows $I(L) = I(L^0)$, and hence θ^* is surjective. (11.2) is thus proved.

(11.3) *Let $e_i^*: H^*(G/L_i^0) \to H^*(G/L^0)$, $\theta_i^*: H^*(G/L_i) \to H^*(G/L_i^0)$ be the homomorphisms induced by the natural maps:*

$$G/L^0 \to G/L_i^0, \quad G/L_i^0 \to G/L_i, \quad (i = 1, 2).$$

Then

$$H^*(G/L_i^0) = \theta_i^*(H^*(G/L_i)) \oplus \ker e_i^*.$$

Proof. Let l be an element of G which normalizes L^0 and L_i^0. We define $\phi(l): G/L^0 \to G/L^0$ to be the homeomorphism $gL \to glL$, and $\phi_i(l): G/L_i^0 \to G/L_i^0$ the homeomorphism $gL_i^0 \to glL_i^0$, $(g \in G)$. Then the diagram

$$\begin{tikzcd} H^*(G/L^0) \arrow[r, "\phi(l)^*"] & H^*(G/L^0) \\ H^*(G/L_i^0) \arrow[u, "e_i^*"] \arrow[r, "\phi_i(l)^*"] & H^*(G/L_i^0) \arrow[u, "e_i^*"'] \end{tikzcd}$$

is commutative. From (11.2), $\phi(l)^*$ is the identity for all $l \in L$. Since L is compact, we see readily from the diagram that $e_i^*(J) = e_i^*(H^*(G/L_i^0))$, where

$$J = \{u: \phi_i(l)^*(u) = u, l \in L\}$$

is the set of all $\phi_i(L)^*$-invariant elements in $H^*(G/L_i^0)$. But $L_i = LL_i^0$, and so J is also the set of all $\phi_i(L_i)$-invariant elements in $H^*(G/L_i^0)$. On the other hand, G/L_i^0 is a covering space of G/L_i with $\phi_i(L_i)$ as its group of covering transformations. Therefore

(11.4) $$J = \theta_i^*(H^*(G/L_i)).$$

From the commutative diagram

$$\begin{array}{ccc} H^*(G/L) & \xrightarrow{\theta^*} & H^*(G/L^0) \\ {\scriptstyle f_i^*}\uparrow & & \uparrow{\scriptstyle e_i^*} \\ H^*(G/L_i) & \xrightarrow{\theta_i^*} & H^*(G/L_i^0) \end{array}$$

we know that, when restricted to J, the homomorphism e_i^* is injective. In other words, $J \cap \ker e_i^* = 0$. Our proposition then follows from (11.4) and the equality $e_i^*(J) = e_i^*(H^*(G/L_i^0))$ which we have established.

(11.5) *If X_2 is orientable, then the homomorphism θ_2^*: $H^*(G/L_2) \to H^*(G/L_2^0)$ induced by the covering map: $G/L_2^0 \to G/L_2$ is bijective. If X_2 is not orientable, then*

$$\mathfrak{P}(X) = (1 + t^{2q+1})\mathfrak{P}(X_2), \qquad \mathfrak{P}(G/L_2^0) = (1 + t^{q+1})\mathfrak{P}(X_2),$$

where "$\mathfrak{P}$" denotes the Poincaré polynomial with t as the indeterminate.

Proof. G/L^0 is an orientable S^q-bundle over G/L_2^0. Therefore we have the Gysin-Chern-Spanier sequence with real coefficients:

$$H^*(G/L_2^0) \xrightarrow{e_2^*} H^*(G/L^0) \xrightarrow{\partial} H^*(G/L_2^0) \xrightarrow{\omega} H^*(G/L_2^0) \xrightarrow{e_2^*} H^*(G/L^0),$$

where e_2^* is degree preserving, while ∂ lowers the degree by q and ω raises the degree by $q+1$. For any graded linear space E, let us denote by $\mathfrak{P}(E)$ the Poincaré polynomial of E with t as the indeterminate. Then from the above exact sequence, we have

$$\mathfrak{P}(G/L^0) = \mathfrak{P}(\operatorname{im} e_2^*) + t^q\mathfrak{P}(\operatorname{im}\partial), \qquad \mathfrak{P}(G/L_2^0) = \mathfrak{P}(\operatorname{im}\partial) + \mathfrak{P}(\operatorname{im}\omega)/t^{q+1},$$
$$\mathfrak{P}(G/L_2^0) = \mathfrak{P}(\operatorname{im}\omega) + \mathfrak{P}(\operatorname{im} e_2^*).$$

Let $J = \theta_2^*(H^*(G/L_2))$ and $K = \ker e_2^* = \operatorname{im}\omega$. Then, we have from (11.3)

$$J = \operatorname{im} e_2^*,\ H^*(G/L_2^0) = J \oplus K,\ \mathfrak{P}(J) = \mathfrak{P}(X_2).$$

Hence

(11.6) $$\mathfrak{P}(G/L^0) = \mathfrak{P}(X_2) + t^q\mathfrak{P}(\operatorname{im}\partial), \qquad \mathfrak{P}(G/L_2^0) = \mathfrak{P}(\operatorname{im}\partial) + \mathfrak{P}(K)/t^{q+1},$$
$$\mathfrak{P}(G/L_2^0) = \mathfrak{P}(K) + \mathfrak{P}(X_2).$$

Case I. Suppose X_2 to be orientable. Then from (3.5) and (11.2), we have $\mathfrak{P}(G/L^0) = \mathfrak{P}(X) = \mathfrak{P}(X_2)(1 + t^q)$. It follows then $\mathfrak{P}(X_2) = \mathfrak{P}(\operatorname{im}\partial)$,

and then $\operatorname{im}\omega = 0$, $\mathfrak{P}(G/L_2^0) = \mathfrak{P}(X_2)$. Since $\theta_2^*: H^*(G/L_2) \to H^*(G/L_2^0)$ is always injective, the equality $\mathfrak{P}(G/L_2^0) = \mathfrak{P}(X_2)$ tells us that θ_2^* is bijective. The first part of (11.5) is thus proved.

Case II. Suppose that X_2 is non-orientable. Let Ω be the characteristic class of the S^q-bundle $\{G/L^0, G/L_2^0, e_2\}$. For each l of L, we define $\phi(l): G/L^0 \to G/L^0$ to be the homeomorphism $gL^0 \to glL^0$ and $\phi_2(l): G/L_2^0 \to G/L_2^0$ to be the homeomorphism $gL_2^0 \to glL_2^0$, $(g \in G)$. Then the pair $\{\phi(l), \phi_2(l)\}$ forms an automorphism of the bundle $\{G/L^0, G/L_2^0, e_2\}$. Therefore $\phi_2(l)^*\Omega = \pm\Omega$, and then Ω^2 is invariant under $\phi_2(l)^*$ for all l of L. Since $L_2 = LL_2^0$, Ω^2 is also left invariant by $\phi_2(L_2)$. The space G/L_2^0 is a covering space of G/L_2 with $\phi_2(L_2)$ as its group of covering transformations. Therefore $\Omega^2 \in \theta_2^*(H^*(G/L_2)) = J$. On the other hand, $e_2^*(\Omega^2) = 0$. (11.3) then tells us that $\Omega^2 = 0$.

We know that the linear map $\omega: H^*(G/L_2^0) \to H^*(G/L_2^0)$ is defined by $\omega(u) = \Omega \cdot u$, $u \in H^*(G/L_2^0)$. It follows that

$$\omega(K) = \omega(\operatorname{im}\omega) = 0, \qquad K = \omega(H^*(G/L_2^0)) = \omega(J + K) = \omega(J).$$

Here G/L_2^0 is an orientable manifold which covers the non-orientable manifold $X_2 = G/L_2$. Therefore [6] $\dim H^*(G/L_2^0) \geqq 2\dim H^*(G/L_2) = 2\dim J$. This together with (11.6) tells us that $\dim K \geqq \dim J$. But ω maps J onto K and so ω restricted to J is injective, whence $\mathfrak{P}(K) = t^{q+1}\mathfrak{P}(J) = t^{q+1}\mathfrak{P}(X_2)$. (11.6) then implies that

$$\mathfrak{P}(G/L_2^0) = (1 + t^{q+1})\mathfrak{P}(X_2), \qquad \mathfrak{P}(G/L^0) = (1 + t^{2q+1})\mathfrak{P}(X_2).$$

This together with (11.2) proves our proposition.

(11.7) *If $p = 1$ and q is even, then either* (a) $X_1 = S^q$, $X_2 = S^1$, $X = S^q \times S^1$, *or* (b) $n = 2q + 3$, $L_2^0 \not\sim 0$ *in* G, $\mathfrak{P}(X_1) = (1 + t^q)(1 + t^{q+1})$, $\mathfrak{P}(X_2) = (1 + t)(1 + t^{q+1})$, $X = S^1 \times X_1$ *and* X_2 *is orientable.*

Proof. Let us consider the restriction map $r: I(L_2^0) \to I(L^0)$. For each l of L, $\operatorname{Ad} l$ acts on both $I(L^0)$ and $I(L_2^0)$, and commutes with the restriction map r. Since L_2^0/L^0 is an even sphere, r is injective. From (11.1), $\operatorname{Ad} l$ acts trivially on $I(L^0)$, so it must act also trivially on $I(L_2^0)$ for all $l \in L$. But $L_2 = LL_2^0$, and hence $\operatorname{Ad} L_2$ leaves $I(L_2^0)$ elementwise fixed. Just as in the proof of (11.2), we know that the homomorphism $\theta_2^*: H^*(G/L_2) \to H^*(G/L_2^0)$ is bijective. X_2 is therefore orientable.

X is a q-sphere bundle over X_2 with $q > 1$. Therefore the homomorphism $f_2^{\#}: \pi_1(X) \to \pi_1(X_2)$ is bijective. (4.3) then implies that $\pi_1(X_1) = 0$,

whence L_1 is connected and X_1 orientable. On account of (3.5), the circle bundle $\{X, X_1, f_1\}$ is trivial. In particular, $X = S^1 \times X_1$.

Now consider the system $\{G, L^0, L_1, L_2^0\}$ of connected groups with $L^0 \subset L_1$, $L^0 \subset L_2^0$, $L_1 \subset G$, $L_2^0 \subset G$. Since both θ^* and θ_2^* are bijective, we have from (3.4) that

$$f_1^*(H^s(G/L_1)) + e_2^*(H^s(G/L_2^0)) = H^s(G/L^0),$$
$$f_1^*(H^s(G/L_1)) \cap e_2^*(H^s(G/L_2^0)) = 0 \qquad (0 < s < n-1).$$

Moreover, $L_1/L^0 = S^1$, $L^0/L_2^0 = S^q$. Using exactly the same argument as in the proof of (8.3), we find that either (i) $G/L_i = S^q$, $G/L_2^0 = S^1$, or (ii) $n = 2q + 3$, $L_2^0 \not\sim 0$ in G, and

$$\mathfrak{P}(G/L_1) = (1 + t^q)(1 + t^{q+1}), \qquad \mathfrak{P}(G/L_2^0) = (1 + t)(1 + t^{q+1}).$$

These together with (11.6) and the orientability of X_2 prove our proposition.

(11.8) *If $p = 1$ and q is odd, then either* (a) *X_1, X_2 are points, or* (b) *X_1, X_2 have the same real homology as some odd-dimensional spheres.*

Proof. Let us consider the restriction map $r_i: I(L_i^0) \to I(L^0)$, $(i = 1, 2)$, Since L_i^0/L^0 is an odd sphere, r_i is surjective. For any l of L, $\mathrm{Ad}\, l$ acts on both $I(L^0)$ and $I(L_i^0)$, and commutes with r_i. By (11.1), $I(L^0)$ is elementwise invariant under $\mathrm{Ad}\, L$. Since L is compact, to each $u \in I(L^0)$, there exists $v_i \in I(L_i^0)$ such that $r_i(v_i) = u$ and v_i is invariant under $\mathrm{Ad}\, L$. But $L_i = LL_i^0$ and so v_i is also $\mathrm{Ad}\, L_i$-invariant. This tells us that v_i belongs to $I(L_i)$. Hence the restriction map $\alpha_i: I(L_i) \to I(L)$ is also surjective. It follows then from (3.3) and the following commutative diagram

$$\begin{tikzcd}
I(L_1) \arrow[r, "\sigma_1"] \arrow[d, "\alpha_1"'] & H^*(X_1) \arrow[d, "f_1^*"] \\
I(L) \arrow[r, "\sigma"] & H^*(X) \\
I(L_2) \arrow[r, "\sigma_2"] \arrow[u, "\alpha_2"] & H^*(X_2) \arrow[u, "f_2^*"']
\end{tikzcd}$$

that $\sigma(I(L))$ contains only element of zero degree. But we know that $I(L) = I(L^0)$ and $\theta^*: H^*(G/L) \to H^*(G/L^0)$ is bijective. Therefore the characteristic subalgebra of the homogeneous space G/L^0 is trivial and then the projection $H^*(G/L^0) \to H^*(G)$ is injective. Since we know that θ^*, f_i^* are all injective, the homomorphism $h^*: H^*(X) \to H^*(G)$, $h_i^*: H^*(X_i)$

$\to H^*(G)$ induced by the projections are all injective $(i=1,2)$. Taking account of (11.5) and using the same argument as used in the proof of (5.2), we find that either $\mathfrak{P}(X_1)=\mathfrak{P}(X_2)=1$, or $\mathfrak{P}(X_1)=1+t^a$, $\mathfrak{P}(X_2)=1+t^b$, where a, b are two odd numbers. (11.8) is thus proved.

(11.9) *Suppose that* $p=1$ *and* q *is odd. Then either* (i) X_1, X_2 *are points,* (ii) $X_1=S^q$, $X_2=S^1$, $X=S^q\times S^1$, (iii) $n=2q+3$, $\mathfrak{P}(X)=(1+t)(1+t^{2q+1})$, $\mathfrak{P}(X_1)=1+t^{2q+1}$, $\mathfrak{P}(X_2)=1+t$, X_2 *is non-orientable,* (iv) $n=5$, $p=q=1$,

$$\mathfrak{P}(X_1)=1+t,\quad \mathfrak{P}(X_2)=1+t^3,\quad \mathfrak{P}(X)=(1+t)(1+t^3),$$

(v) $n=4$, $p=q=1$, $\mathfrak{P}(X_1)=\mathfrak{P}(X_2)=1$, $\mathfrak{P}(X)=1+t^3$, *both* X_1 *and* X_2 *are non-orientable, or* (vi) $n=7$, $p=q=1$, $\mathfrak{P}(X)=(1+t^3)^2$, $\mathfrak{P}(X_1)=\mathfrak{P}(X_2)=(1+t^3)$, *both* X_1 *and* X_2 *are non-orientable.*

Proof. Let us divide our discussions into four cases according to the orientability of X_1 and X_2.

Case a. Both X_1 and X_2 are orientable. From (11.5) and (11.8), we know that either X_1 and X_2 are points, or X_1 and X_2 are homology spheres with real coefficients of odd dimension. In the second alternative, it follows from (3.4) and the orientability of X_1, X_2 that $\dim X_1+\dim X_2=\dim X$, and then (4.8) tells us that $X_1=S^q$, $X_2=S^1$, $X=S^q\times S^1$. Thus (11.9) is proved for this case.

Case b. X_1 is orientable but X_2 is not. From (3.4), (3.5) and (11.5), we have

$$\mathfrak{P}(X)=(1+t)\mathfrak{P}(X_1)=(1+t^{2q+1})\mathfrak{P}(X_2)=\mathfrak{P}(X_1)+\mathfrak{P}(X_2)+t^{n-1}-1,$$

whence

$$\mathfrak{P}(X_1)=(1+t^{2q+1})(1-t^{n-1})/(1-t^{2q+2}),$$
$$\mathfrak{P}(X_2)=(1+t)(1-t^{n-1})/(1-t^{2q+2}).$$

By (11.8), X_1 and X_2 are homology spheres, and therefore,

$$n=2q+3,\ \mathfrak{P}(X_1)=1+t^{2q+1},\ \mathfrak{P}(X_2)=1+t.\ \mathfrak{P}(X)=(1+t)(1+t^{2q+1}).$$

(11.9) is hence valid in this case.

Case c. X_2 is orientable but X_1 is not. Suppose $q>1$. Then $f_2^{\#}$: $\pi_1(X)\to\pi_1(X_2)$ is bijective and then, by (4.3), X_1 is simply-connected. This contradicts the non-orientability of X_1, and therefore, we must have $q=1$.

Thus this case is reduced to a particular case of Case b with X_1 and X_2 interchanged. We have then

$$\mathfrak{P}(X_1) = 1 + t,\ \mathfrak{P}(X_2) = 1 + t^3,\ \mathfrak{P}(X) = (1+t)(1+t^3),\ n = 5.$$

Case d. Both X_1 and X_2 are non-orientable. From (4.3), we must have $q = 1$, and then, from (11.5) and (3.4)

$$\mathfrak{P}(X) = (1+t^3)\mathfrak{P}(X_1) = (1+t^3)\mathfrak{P}(X_2) = \mathfrak{P}(X_1) + \mathfrak{P}(X_2) + t^{n-1} - 1,$$

whence

$$\mathfrak{P}(X_1) = \mathfrak{P}(X_2) = (1 - t^{n-1})/(1 - t^3).$$

But from (11.8), X_1 is a homology sphere and therefore either $n = 4$, $\mathfrak{P}(X) = 1 + t^3$, $\mathfrak{P}(X_1) = \mathfrak{P}(X_2) = 1$, or $n = 7$, $\mathfrak{P}(X) = (1+t^3)^2$, $\mathfrak{P}(X_1) = \mathfrak{P}(X_2) = 1 + t^3$. This completes the proof.

12. Determination of G when $p = 1$. Throughout this section, we always assume that $p = 1$. Let us first study the case by which q is even. Then L_1 is connected, X_1 simply-connected and $X = S^1 \times X_1$. Neglecting the case $X_1 = S^q$, $X_2 = S^1$, we suppose that Case b in (11.7) happens. Then $L_2{}^0 \not\sim 0$ in G and

$$n = 2q + 3,\ \mathfrak{P}(X_1) = (1+t^q)(1+t^{q+1}),\ \mathfrak{P}(X_2) = (1+t)(1+t^{q+1}).$$

Using the same kind of arguments as (but much simpler than) those used in §§ 8, 9, 10, we find that, when $n > 15$, then up to a finite covering,

$$\begin{gathered} G = O(2) \times O(q+2),\ L^0 = O(q),\ L_1 = L_1{}^0 = T \cdot O(q),\ L_2{}^0 = O(q+1), \\ T \not\subset O(2),\ T \cap O(q+2) = e, \end{gathered} \tag{12.1}$$

where T is a circle group centralizing L^0, $O(r)$ denotes the rotation group in r variables, and the imbeddings $O(q) \subset O(q+1)$, $O(q+1) \subset O(q+2)$ are in the canonical manner. Here, unlike the situation in § 10, the subgroups L, L_1, L_2 are not uniquely determined by these properties. We have to discuss the components of L and the position of T in G. For simplicity, let us assume

$$O(2) \cap L = e. \tag{12.2}$$

No generality is lost here since $O(2) \cap L$ leaves every point of S^n fixed. We note that after factoring out $O(2) \cap L$ from $O(2)$, G remains the same form as direct product.

(12.3) *Under the above assumptions, L and L_2 have at most two connected components.*

Proof. Since L_2/L is an even sphere, L_2 and L have the same number of connected components, and hence it is sufficient for us to discuss $L_2/L_2{}^0$. Let $\xi\colon G \to O(q+2)$ be the projection. Then $\xi(L_2)$ belongs to the normalizer of $L_2{}^0$ in $O(q+2)$. But from (12.1), $L_2{}^0$ is an $O(q+1)$ imbedded in $O(q+2)$ in the canonical manner, and so $\xi(L_2)/L_2{}^0$ is at most of order 2. Now let us consider the intersection $O(2)\cap L_2$ of L_2 and the kernel of ξ. Since $L \subset L_1$, there exist elements $a_1, a_2, \cdots, a_s \in T$ such that $L = \bigcup_i a_i L^0$, whence $L_2 = \bigcup_i a_i L_2{}^0$. Suppose $b \in O(2)\cap L_2$. There is an element c of $L_2{}^0$ such that $b = a_i c$ for a certain i. Both b and a_i centralize L^0, so does c. Since q is even, the centralizer of $O(q)$ in $O(q+1)$ belongs to $O(q)$. Therefore, $c \in O(q) = L^0$ and then $b \in L$. On account of (12.2), b must be the identity. In other words, $O(2)\cap L_2 = e$, whence $L_2/L_2{}^0 \approx \xi(L_2)/L_2{}^0$ is at most of order 2. (12.3) is thus proved.

(12.4) *The assumptions are as above. Let m be the order of the cyclic group $O(2)\cap T$, $e_1\colon G/L^0 \to G/L_1$ the natural map, and Y the submanifold $O(2)\times O(q+1)/O(q)$ of G/L^0. Denote by $[u]$ the integral homology class of the fundamental cycle of Y. Then e_1 carries $[u]$ into $2m$ times a generator of the $(q+1)$-th homology group $H_{q+1}(G/L_1, Z)$ with integer coefficients.*

Proof. Let C be the identity component of the centralizer of L^0 in $O(q+2)$. From (12.1), C is a circle group. Since $T\cap O(q+2) = e$, and $O(2)\cap T$ is of order m, we can write

$$T = \{(a, \beta(a)^m) : a \in O(2)\},$$

where $\beta\colon\colon O(2) \to C$ is one of the two isomorphisms of $O(2)$ on C. Now let us consider the following two maps

$$G/L_1 \to O(q+2)/O(q), \qquad O(q+2)/O(q) \to O(q+2)/O(q+1)$$

defined, respectively, by

$$agL_1 \to g\beta(a)^{-m}O(q), \qquad gO(q) \to gO(q+1),$$
$$a \in O(2), \qquad g \in O(q+2).$$

The first map is bijective while the second is the projection in the tangent bundle over an S^{q+1}. Since $q+1$ is odd, the composite map

$$j\colon G/L_1 \to O(q+2)/O(q+1)$$

of the above two maps induces an isomorphism between the $(q+1)$-th integral homology groups $H_{q+1}(G/L_1{}^0)$ and $H_{q+1}(O(q+2)/O(q+1))$. Thus, in order

to prove our proposition, it suffices to prove that the map

$$j \circ e_1 \colon Y \to O(q+2)/O(q+1)$$

is of degree $2m$. For this purpose, let us note that $O(q+2)/O(q+1) = S^{q+1}$, $Y = O(2) \times O(q+1)/O(q) = S^1 \times S^q$, and the map $j \circ e_1$ is given by $agO(q) \to g\beta(a)^{-m}O(q+1)$, $a \in O(2)$, $g \in O(q+1)$. By using orthogonal matrices, we can, in a very natural manner, identify Y with the subset

$$\{(\cos\theta, -\sin\theta, x_1, x_2, \cdots, x_{q+1}) : 0 \leqq \theta \leqq 2\pi, \sum x_i^2 = 1\}$$

of R^{q+3} and identify S^{q+1} with the subset $\{(y_0, y_1, \cdots, y_{q+1}) : \sum y_i^2 = 1\}$ of R^{q+2} such that $j \circ e_1$ is given by

$$y_0 = \cos m\theta, \quad y_i = -x_i \sin m\theta, \qquad i = 1, 2, \cdots, q+1.$$

It is direct to verify that $j \circ e_1$ is of degree $2m$ either by local degree argument or by using differential forms. (12.4) is thus proved.

(12.5) *The assumptions are as above. Then* $O(2) \cap T = e$, *and* L, L_2 *have exactly two connected components.*

Proof. Let Z_k denote the additive group of integers mod k, $k \neq 1$. Then the commutative diagram

$$\begin{tikzcd} & G/L^0 \arrow[r, "e_2"] \arrow[dl, "e_1"'] \arrow[dd, "\theta"] & G/L_2^0 \arrow[dd, "\theta_2"] \\ X_1 & & \\ & X \arrow[ul, "f_1"] \arrow[r, "f_2"] & X_2 \end{tikzcd}$$

induces a commutative diagram of the $(q+1)$-th homology groups with Z_k as the coefficient group

$$\begin{tikzcd} & H_{q+1}(G/L^0, Z_k) \arrow[r, "e_2"] \arrow[dl, "e_1"'] \arrow[dd, "\theta"] & H_{q+1}(G/L_2^0, Z_k) \arrow[dd, "\theta_2"] \\ H_{q+1}(X_1, Z_k) & & \\ & H_{q+1}(X, Z_k) \arrow[ul, "f_1"] \arrow[r, "f_2"] & H_{q+1}(X_2, Z_k). \end{tikzcd}$$

$Y = O(2) \times O(q+1)/O(q)$ is a submanifold of $G/L^0 = O(2) \times O(q+2)/O(q)$. Let u be the fundamental cycle of Y, and $[u]_k$ the element of $H_{q+1}(G/L^0, Z_k)$ which contains u. Since $q+1$ is odd, we know from the familiar properties

of the Stiefel manifolds that $[u]_k \neq 0$ for any $k \neq 1$. The map e_2 carries Y into a circle, and hence

$$e_2([u]_k) = 0, \qquad f_2\theta([u]_k) = \theta_2 e_2([u]_k) = 0.$$

From (12.4), $e_1([u]_{2m}) = 0$, and then by the above commutative diagram, $f_1\theta([u]_{2m}) = 0$. Therefore $\theta([u]_{2m})$ belongs to both the kernel of f_1 and the kernel of f_2. Since $0 < q+1 < n-1$, it follows from the dual of (3.4) that $\theta([u]_{2m}) = 0$.

If L is connected, then θ becomes the identity map, and then $[u]_{2m} = \theta([u]_{2m}) = 0$ which is impossible. Therefore by (12.3), both L and L_2 have exactly two connected components. Now we know that $\theta\colon G/L^0 \to X$ is a two-fold covering map, and moreover, from (11.2) that θ induces an isomorphism on the homology groups with *real* coefficients. Therefore, $\theta([u]_k) \neq 0$ for $k > 2$. This implies that $2m \leqq 2$, or in other words, $m = 1$, $O(2) \cap T = e$. (12.5) is thus proved.

(12.6) *The assumptions are as above. Then $(O(2) \times O(q+2))/I_2$ acts effectively on S^n, where I_2 is the only central subgroup of order 2 which belongs to neither $O(2)$ nor $O(q+2)$. Moreover, the triple $\{L, L_1, L_2\}$ of subgroups is uniquely determined up to an automorphism of G.*

Proof. From (12.1), the pair $\{L^0, L_2{}^0\}$ of subgroups of $G = O(2) \times O(q+2)$ is unique up to an automorphism of $O(q+2)$. From (12.1) and (12.5), T is a circle group centralizing L^0 and $T \cap O(2) = T \cap O(q+2) = e$. Therefore T is unique up to an automorphism of $O(2)$. Now let us see that T, L^0, $L_2{}^0$ completely determine the triple $\{L, L_1, L_2\}$. Since $L_1 = T \cdot L^0$, L_1 is settled. Noting that L/L^0 is of order 2, $L \subset L_1$ and $L \cap O(2) = e$, we find immediately that L can only be $L_2 L$. Therefore, we have

$$L_1 = T \cdot L^0,\; L = I_2 L^0,\; L_2 = L L_2{}^0 = I_2 L_2{}^0.$$

Thus, up to an automorphism of $O(2) \cap O(q+2)$, the system $\{T, L^0, L_2{}^0\}$ and hence the system $\{L, L_1, L_2\}$ is unique. This establishes the second part of (12.6). The first part of (12.6) follows from the fact that I_2 is the only normal subgroup of G contained in L. The proof of our proposition is completed.

As a direct consequence of (11.5), (12.1) and (12.6), we have

(12.7) *Let G be a compact, connected differentiable transformation group of S^n with $(n-1)$-dimensional orbits. Suppose that $n > 15$, $p = 1$, q is even, and the special orbits are not spheres. Then $n = 2q+3$, X_1 is the Stiefel manifold $V_{q+2,2}$, $X = S^1 \times X_1$, and X_2 is an orientable S^{q+1}-bundle*

over S^1. *Moreover, if* G *is effective, then* $G \approx (O(2) \times O(q+2))/I_2$, *where* I_2 *is the central subgroup of* $O(2) \times O(q+2)$ *belonging to neither of the direct factors. The system* $\{L, L_1, L_2\}$ *of subgroups is unique up to an automorphism of* G.

Now let us come to the case by which $p = 1$ and q is odd. Taking account of (11.9) and using similar arguments as in the case of even q, we can prove

(12.8) *Let* G *be a compact, connected differentiable transformation group of* S^n *with an* $(n-1)$*-dimensional orbit. Suppose that* $p = 1$, q *is odd,* $n > 13$, *and the special orbits are neither both points nor* (S^q, S^1). *Then* $n = 2q + 3$, X_1 *is the Stiefel manifold* $V_{2q+2,2}$, $X = S^1 \times X_1$, *and* X_2 *is the* $(q+2)$*-dimensional generalized Klein bottle. If* G *is, moreover, effective, then* $G \approx O(2) \times O(q+2)$ *and the triple* $\{L, L_1, L_2\}$ *of subgroups is unique up to an automorphism of* G.

13. Summary of the results. For clarity, let us recall some of the relevant definitions and summarize the results which have been obtained. G is a compact, connected and differentiable transformation group of S^n with an $(n-1)$-dimensional orbit. X denotes a general orbit, X_1 and X_2 denote the two special orbits, and $p = \dim X - \dim X_1$, $q = \dim X - \dim X_2$. L, L_1, L_2 are the isotropic subgroups of G at points a, a_1, a_2 on X, X_1, X_2 respectively, where a, a_1, a_2 are so chosen that $L \subset L_1$, $L \subset L_2$. From (5.2), (7.4), (8.3), (10.1), (10.4), (11.7), (11.9), (12.7) and (12.8), we have

THEOREM I. *If* n *is even and different from* 4, *then either* (i) X_1, X_2 *are single points,* $X = S^{n-1}$, *or* (ii) $X_1 = S^q$, $X_2 = S^p$, $X = S^q \times S^p$.

THEOREM II. *Suppose that* G *is effective,* $n > 31$ *is odd, and* $q \leqq p$ *(up to a change of notation, we can always assume* $q \leqq p$*). Then there are only five possibilities:*

(i) X_1, X_2 *are single points,* $X = S^{n-1}$;

(ii) $X_1 = S^q$, $X_2 = S^p$, $X = S^q \times S^p$;

(iii) $q = 4$, $p = 4r - 5$, $n - 1 = 2(p+4)$, X_1 *is an* S^{p+4}*-bundle over* S^4, $X = S^4 \times X_2$, X_2 *is the quaternion Stiefel manifold* $Sp(r)/Sp(r-2)$, $G \approx (Sp(r) \times Sp(2))/I_2$, *where* I_2 *is the central subgroup of order* 2 *belonging to neither* $Sp(r)$ *nor* $Sp(2)$;

(iv) $q = 2$, $p = 2r - 3$, $n - 1 = 2(p+2)$, X_1 *is an* S^{p+2}*-bundle over* S^2, $X = S^2 \times X_2$, X_2 *is the complex Stiefel manifold* $SU(r)/SU(r-2)$, *and* G *is isomorphic with either* $(SU(r) \times SU(2))/I$ *or* $(U(r) \times U(2))/T$, *where*

I and T are the maximal central subgroups which intersect each direct factor only at the indentity;

(v) $q=1$, $n=2p+3$, *X_1 is an S^{p+1}-bundle over S^1, X_2 is the real Stiefel manifold $O(p+2)/O(p)$, $X=S^1\times X_2$, and G is isomorphic with $O(p+2)\times O(2)$ or $(O(p+2)\times O(2))/I_2$ according as p is odd or even, where I_2 denotes the central subgroup of order 2 belonging to neither of the direct factors. Moreover, in Cases* (iii), (iv) *and* (v), *the triple $\{L, L_1, L_2\}$ of subgroups is unique up to an automorphism of G.*

Concerning the orbits of G, we know from the above theorems that there are five possibilities. Now let us see that all of them are actually realized by subgroups of $O(n+1)$ acting on the unit sphere in R^{n+1}.

(i) Let G be any closed and connected subgroup of $O(n)$ transitive on the unit sphere S^{n-1} of R^n. By imbedding $O(n)$ in $O(n+1)$ in the canonical manner, the group G acts on the unit sphere S^n of R^{n+1} with an $(n-1)$-dimensional orbit and with two single points as the special orbits.

(ii) Let G', G'' be connected closed subgroups of $O(q+1)$, $O(p+1)$ transitive on the unit spheres S^q, S^p of R^{q+1}, R^{p+1} respectively. Since $p+q=n-1$, we can identify the direct sum $R^{q+1}\otimes R^{p+1}$ with R^{n+1}, and thus G', G'' are imbedded in $O(n+1)$ as subgroups. The direct product $G'\times G''$ acts on the unit sphere of R^{n+1} with an $(n-1)$-dimensional orbit and with S^q, S^p as the two special orbits.

(iii) Let $Sp(r)$ be the group of all symplectic transformations of the r-dimensional right quaternion vector space Q^r and $Sp(2)$ that of the left quaternion plane Q^2. Then $Sp(r)\times Sp(2)$ acts naturally (not effectively) on the tensor product $Q^r\otimes Q^2$ which is the real vector space R^{8r}. Let $Sp(r)\otimes Sp(2)$ denote the group of automorphisms of R^{8r} so obtained. Then $Sp(r)\otimes Sp(2)$ acts on the unit sphere S^n of R^{8r}, $n=8r-1$, with an $(n-1)$-dimensional orbit. The special orbits consist of a quaternion Stiefel manifold and an S^{4r-1}-bundle over S^4. We note that $Sp(r)\times Sp(2)$ is a two-fold covering group of $Sp(r)\otimes Sp(2)$.

(iv) The tensor products $SU(r)\otimes SU(2)$ and $U(r)\otimes U(2)$ are groups of unitary transformations of the complex vector space C^{2r}, and hence are transformation groups of the unit sphere S^n of C^{2r}, where $n=4r-1$. Both of these groups have an $(n-1)$-dimensional orbit. The two special orbits consists of a complex Stiefel manifold and an S^{p+2}-bundle over S^2. It is noted that $SU(r)\otimes SU(2)\approx(SU(r)\times SU(2))/I$, $U(r)\otimes U(2)\approx(U(r)\times U(2))/T$,

where I is the identity for odd r and is of order 2 for even r, and T is a circle group.

(v) The tensor product $G = O(p+2) \otimes O(2)$ which acts on R^{2p+4} orthogonally gives a realization of the possibility (v) in Theorem II.

14. Orthogonality of G. Let K be a transformation group of a space M. By a linear representation of $\{K, M\}$ of degree m, we mean a pair $\{\alpha, \phi\}$, where ϕ is a continuous map of M into the euclidean space R^m and $\alpha \colon K \to GL(R, m)$ is a homomorphism such that

$$\alpha(k)(\phi(x)) = \phi(k(x)), \; k \in K, \; x \in M.$$

Now suppose that M is an n-sphere. We say that K acts on M orthogonally if there exists a linear representation $\{\alpha, \phi\}$ of degree $n+1$ such that ϕ carries M homeomorphically onto the unit sphere in R^{n+1}. It is to be noted that we do not assume K to be effective and hence α may not be an isomorphism.

It is the aim of the present section to show that if $n > 31$, then G acts on S^n orthogonally. For this purpose, we find it convenient to write the difference set $S^n - (X_1 \cup X_2)$ in the form of a topological product. Let us recall that, in our discussions, we take a cross section $\widehat{a_1 a_2}$ of the orbits of G. This cross section is an arc with a_1, a_2 as the end points which has the properties that the isotropc subgroups at any two inner points coincide, and $G(a_1) = X_1$, $G(a) = X_2$ are the two special orbits. Let us parametrize this arc by t in such

$$\widehat{a_1 a_2} = \{s(t) : -1 \leqq t \leqq 1\}, \; s(-1) = a_1, \; s(1) = a_2,$$

and $G(s(0))$ is our orbit X. This point $s(0)$ coincides with the point "a" which we used in § 3. There is a natural homeomorphism between the product $I_0 \times X$ and the difference set $S^n - (X_1 \cup X_2)$, where $I_0 = \{t \colon -1 < t < 1\}$. In fact, for any (t, x) with $x \in X$, $-1 < t < 1$, there exists $g \in G$ such that $g(s(0)) = x$. Since all the isotropic subgroups at $s(t)$ coincide when $-1 < t < 1$, the point $g(s(t))$ is independent of the choice of g. Thus we get a map $(t, x) \to g(a(t))$ which can be easily verified to be continuous and bijective. From now on, we identify $I_0 \times X$ with $S^n - (X_1 \cup X_2)$ by means of this map. It is evident that

$$g(t, x) = (t, g(x)), \; -1 < t < 1, \; g \in G, \; x \in X.$$

THEOREM III. *Let G be a compact, connected differentiable transformation group of S^n with an $(n-1)$-dimensional orbit. If the special orbits are either* (i) *both single points, or* (ii) *spheres S^q, S^p with $p + q = n - 1$, then G acts on S^n orthogonally.*

Proof. Case (i) has been proved by Poncet [17]. It suffices for us to study the case by which $X_1 = S^q$, $X_2 = S^p$ with $p + q = n - 1$. Since G is compact and connected and acts on X_i transiticely, it must act on X_i orthogonally $(i = 1, 2)$ [17]. There exist then a homomorphism α_1: $G \to GL(R, q+1)$ and a map $\phi_1\colon X_1 \to R^{q+1}$ such that $\alpha_1(g)\phi_1(x_1) = \phi_1(g(x_1))$, $x_i \in X_i$, $g \in G$, and ϕ_1 carries X_1 homeomorphically onto the unit sphere in R^{q+1}. Similarly, we have a homomorphism $\alpha_2\colon G \to GL(R, p+1)$ and a map $\phi_2\colon X_2 \to R^{p+1}$ such that $\alpha_2(g)\phi_2(x_2) = \phi_2(g(x_2))$, $x_2 \in X_2$, $g \in G$ and ϕ_2 carries X_2 homeomorphically onto the unit sphere of R^{p+1}.

Let us define $\phi\colon S^n \to R^{q+1} \oplus R^{p+1} = R^{n+1}$ as follows:

$$\begin{cases} \phi(x_1) = (\phi_1(x_1), 0), \ x_1 \in X_1, \\ \phi(x_2) = (0, \phi_2(x_2)), \ x_2 \in X_2. \\ \phi(t,x) = ((t+1)\phi_1 f_1(x), (t-1)\phi_2 f_2(x))/[(t-1)^2 + (t+1)^2]^{\frac{1}{2}}, \\ \qquad\qquad (t,x) \in I_0 \times X = S^n - (X_1 \cup X_2). \end{cases}$$

Here, $\phi_1 f_1(x)$ and $\phi_2 f_2(x)$ are points in euclidean spaces, and so the multiplication by real numbers has a meaning. It is evident that ϕ is continuous and $\phi(S^n)$ belongs to the unit sphere of R^{n+1}. From (4.8), ϕ is injective and hence ϕ carries S^n homeomorphically onto the unit sphere of R^{n+1}. Let us consider R^{q+1}, R^{p+1} to be subspaces of R^{n+1} in the natural manners. Thus $GL(R, q+1)$, $GL(R, p+1)$ are imbedded in $GL(R, n+1)$ as subgroups. Defining $\alpha\colon G \to GL(R, n+1)$ by $\alpha(g) = \alpha_1(g)\alpha_2(g)$, and taking account of the facts that, for $i = 1, 2$.

$$\alpha_i(g)(\phi_i(f_i(x))) = \phi_i(g(f_i(x))), f_i(g(x)) = g(f_i(x)), \ g \in G, \ x \in X,$$

we find immediately that the pair $\{\alpha, \phi\}$ forms a linear representation of $\{G, S^n\}$ of degree $n+1$. Hence G acts on S^n orthogonally. Theorem III is thus proved.

To discuss the orthogonality of G when the special orbits are neither points nor spheres, we need the following proposition:

(14.1) *Let G be a compact, connected differentiable group acting on S^n with an $(n-1)$-dimensional orbit, and L, L_1, L_2 have the same meaning as before. Suppose that G also acts on another sphere S^m with an $(n-1)$-dimensional orbit, and that the corresponding isotropic subgroups be L', L_1', L_2'. If there exists an automorphism θ of G carrying $\{L, L_1, L_2\}$ into $\{L', L_1', L_2'\}$, then $n = m$ and the two transformation groups $\{G, S^n\}$, $\{G, S^m\}$ are equivalent.*

Proof. Since $n-1=\dim G-\dim L=\dim G-\dim L'=m-1$, we have $n=m$. Let $\widehat{a_1a_2}$ ($\widehat{a_1'a_2'}$) be the cross section in S^n (S^m) of orbits based on which L, L_1, L_2 (L', L_1', L_2') are defined. We write

$$\widehat{a_1a_2}=\{s(t):-1\leqq t\leqq 1\}, \qquad \widehat{a_1',a_2'}=\{s'(t):-1\leqq t\leqq \}$$

such that

$$s(-1)=a_1, \quad s(1)=a_2, \quad s'(-1)=a_1', \quad s'(1)=a_2'.$$

This facilitates us to construct a homeomorphism of S^n onto S^m. In fact, let $x\in S^n$. There exists $g\in G$ and t, $-1\leqq t\leqq 1$, such that $x=g(s(t))$. Here t is uniquely determined by x while g is not. Suppose that $x=h(s(t))$, $h\in G$; then $g^{-1}h$ belongs to L_1, L_2 or L according as $t=-1$, $t=1$ or $-1<t<1$. Let us define ϕ: $S^n\to S^m$ by $\phi(x)=\theta(g)(s'(t))$, where g and t satisfy $x=g(s(t))$. Since $\theta(L)=L'$, $\theta(L_1)=L_1'$, $\theta(L_2)=L_2'$, L_1' leaves a_1' invariant, L_2' leaves a_2' invariant and L' leaves $\widehat{a_1'a_2'}$ pointwise invariant, it follows that ϕ is independent of the choice of g and hence well-defined. We can easily verify that ϕ is continuous and injective. Therefore ϕ maps S^n homeomorphically onto S^m as they have the same dimension. Moreover, from the definition of ϕ, we have $\phi(g(x))=\theta(g)(\phi(x))$, $x\in S^n$, $g\in G$. The two transformation groups $\{G,S^n\}$ and $\{G,S^m\}$ are equivalent.

THEOREM IV. *Suppose that n is either even and different from 4, or odd and greater than 31. Then any compact and connected differentiable transformation group of S^n with an $(n-1)$-dimensional orbit is orthogonal.*

Proof. If the special orbits are both points or both spheres, then this is contained in Theorem III. Therefore, it is sufficient to discuss the Cases (iii), (iv), and (v) in Theorem II. From (14.1), the transformation group $\{G,S^n\}$ is determined by the system $\{G,L,L_1,L_2\}$, and from Theorem II, the system $\{G,L,L_1,L_2\}$ is completely determined by the topological group G. Hence, for a given n, there is at most one transformation group G (up to equivalence) realizing Case (iii) or Case (v), while there are at most two transformation groups realizing Case (iv). It follows that the orthogonal models which we have constructed in § 13 exhaust all the transformation groups G for Cases (iii), (iv) and (v). Theorem IV is thus proved.

15. Some exceptional cases. In § 7, we have classified all the orbits of G when both p and q are even. By using (14.1), the transformation groups $\{G,S^n\}$, in this case, can be quite easily determined. We shall give these

groups explicitly in order to see some examples different from the five main classes mentioned in Theorem IV.

THEOREM V. *Let G be a compact, connected differentiable transformation group of S^n with an $(n-1)$-dimensional orbit such that the differences of the dimensions of the orbits are always even. Then n is odd and G acts on S^n orthogonally. If the special orbits are neither both single points nor both spheres, then the transformation group G is either the linear adjoint group of a simple group of rank 2, or the irreducible representation of C_3 of degree 14, or the irreducible representation of the group F_4 of degree 26.*

Proof. The fact that n is odd follows directly from (7.4). On account of Theorem I, we can assume that the special orbits are neither both points nor both spheres. Therefore, by (7.4), there are only five possibilities: (a) $G=A_2$, $L=T^2$, L_1 and L_2 are locally isomorphic with $T\times A_1$; (b) $G=C_2$, $L=T^2$, L_1 and L_2 are locally isomorphic with $T\times A_1$, (c) $G=G_2$, $L=T^2$, L_1 and L_2 are locally isomorphic with $T\times A_1$, (d) $G=C_3$, L is locally isomorphic with $C_1\times C_1\times C_1$, L_1 and L_2 are locally isomorphic with $C_1\times C_2$, (e) $G=F_4$, $L=D_4$, L_1 and L_2 are of Cartan class B_4. If G is, moreover, effective, than G has trivial center. We can verify directly that the compact linear groups given in our theorem give orthogonal realizations of these five cases. Hence, to prove our theorem, it suffices to prove that, for a given G, the action of G of S^n is unique. On account of (14.1), it is sufficient to show that, for each G, the triple $\{L, L_1, L_2\}$ of subgroups is unique up to an automorphism of G. Let us discuss the five cases separtely. We shall give the detail for $G=C_3$. As for the rest, only indications will be provided.

1. $G=C_3$. Then L is locally isomorphic with $C_1\times C_1\times C_1$, and L_1, L_2 with $C_1\times C_2$. Let $\{L', L_1', L_2'\}$ be another such triple of subgroups of G. We shall show that there exists an automorphism of G carrying L, L_1, L_2 into L', L_1', L_2' respectively. Since both L, L' are connected and locally isomorphic with $C_1\times C_1\times C_1$, they must be conjugate [14], and thus we can assume $L=L'$. Take a maximal toral subgroup H of L and consider it Lie algebra $\boldsymbol{H}$. Evidently $\boldsymbol{H}$ is a Cartan subalgebra of the Lie algebra $\boldsymbol{G}$ of G. We can choose coordinates τ_1, τ_2, τ_3 of $\boldsymbol{H}$ such that the angular parameters are

$$\pm 2\tau_i, \pm \tau_i \pm \tau_j \qquad (i, j=1,2,3).$$

For each angular parameter ρ, let us denote by Y_ρ, Z_ρ the quasi-root vectors [14] corresponding to $\pm\rho$. It is known that any subalgebra containing $\boldsymbol{H}$ is spanned by $\boldsymbol{H}$ and some quasi-root vectors. We see immediately that the

Lie algebra $\boldsymbol{L}$ of L is spanned by $\boldsymbol{H}$ and $Y_{2\tau_i}$, $Z_{2\tau_i}$, $i=1,2,3$. A direct computation tells us that there are only three subalgebras $\boldsymbol{J}_1$, $\boldsymbol{J}_2$, $\boldsymbol{J}_3$ of $\boldsymbol{G}$ which contain $\boldsymbol{L}$ and are of type $C_1\times C_2$, namely

$$\boldsymbol{J}_1=\{\boldsymbol{L}, Y_{\tau_1+\tau_2}, Z_{\tau_1+\tau_2}, Y_{\tau_1-\tau_2}, Z_{\tau_1-\tau_2}\},$$
$$\boldsymbol{J}_2=\{\boldsymbol{L}, Y_{\tau_1+\tau_3}, Z_{\tau_1+\tau_3}, Y_{\tau_1-\tau_3}, Z_{\tau_1-\tau_3}\},$$
$$\boldsymbol{J}_3=\{\boldsymbol{L}, Y_{\tau_2+\tau_3}, Z_{\tau_2+\tau_3}, Y_{\tau_2-\tau_3}, Z_{\tau_2-\tau_3}\}.$$

We know that the Weyl group consists of transformations of the form

$$\tau_j=\epsilon_j\tau_{\sigma(j)}, \qquad \epsilon_j=\pm 1, \qquad\qquad j=1,2,3,$$

where σ denotes a permutation of the indices. It follows that, given any two pairs of the algebras $\boldsymbol{J}_1$, $\boldsymbol{J}_2$, $\boldsymbol{J}_3$, there exists an automorphism of $\boldsymbol{G}$ leaving $\boldsymbol{L}$ invariant and carrying one pair to the other. In other words, there is essentially only one pair of subalgebras of the type $C_1\times C_2$ which contains $\boldsymbol{L}$. Since L_1, L_2, L_1', L_2' are connected and $L_1\neq L_2$, $L_1'\neq L_2'$ on account of (6.2), we can find an automorphism of G leaving L invariant and carrying L_1, L_2 into L_1', L_2' respectively. Our theorem is thus proved in this case.

2. $G=F_4$. The proof is the same as before. We note the following two properties of F_4 which can be readily verified: (1) up to an automorphism of F_4, there is only one connected subgroup of type D_4; (2) given any two pairs of connected subgroups of type B_4 which contains a fixed subgroup L of type D_4; there exists an inner automorphism of F_4 leaving L invariant and carrying one pair to the other.

3. $G=A_2$. The proof is the same as the case $G=C_3$.

4. $G=G_2$. L is a two-dimensional toral group and L_1, L_2 are locally isomorphic with $T\times A_1$. There are six connected subgroups which contain L and are locally isomorphic with $T\times A_1$, and hence 15 pairs of such subgroups. Up to automorphisms of G leaving L invariant, there are only four different pairs. Among these four pairs, only one pair satisfies both (6.2) and (7.5). Therefore, the triple $\{L, L_1, L_2\}$ of subgroups of G_2 is unique up to automorphisms of G_2.

5. $G=C_2$. The proof is the same as the preceding one.

REFERENCES.

[1] A. Borel, "Le plan projectif des octaves et les sphères comme espaces homogènes," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 230 (1930), pp. 1378-1380.

[2] ———, "Les bouts des espaces homogènes de groupes de Lie," *Annals of Mathematics*, vol. 58 (1953), pp. 443-457.

[3] ——— et J. De Siebenthal, "Les sous-groupes fermés de rang maximum des groupes de Lie clos," *Commentarii Mathematici Helvetici*, vol. 23 (1949), pp. 200-221.

[4] H. Cartan, "La transgression dans un groupe de Lie et dans un espace fibré principal," *Colloque de topologie*, Bruxelles, 1950, Liège et Paris, 1951, 57-71.

[5] S. S. Chern and E. Spanier, "The homology structure of sphere bundles," *Proceedings of the National Academy of Sciences of the U. S. A.*, vol. 36 (1950), pp. 248-255.

[6] B. Eckmann, "Covering spaces and Betti numbers," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 95-101.

[7] J. Leray, "L'homologie d'un espace fibré dont la fibre est connexe," *Journal de Mathématiques pures et appliquées*, vol. 29 (1950), pp. 169-213.

[8] ———, "Sur l'homologie des groupes de Lie, des espaces homogènes et des espaces fibrés principaux," *Colloque de topologie algebrique*, Bruxelles, 1950, Liège et Paris, 1951, 101-115.

[9] D. Montgomery and H. Samuelson, "Transformation groups of spheres," *Annals of Mathematics*, vol. 44 (1943), pp. 454-470.

[10] ——— and L. Zippin, "A clas of transformation groups in E_n," *American Journal of Mathematics*, vol. 63 (1941), pp. 1-8.

[11] ———, *Topological transformation groups*, New York, 1955.

[12] P. Olum, "Non-abelian cohomology and van Kampen's theorem," *Annals of Mathematics*, vol. 68 (1958), pp. 658-668.

[13] H. Samelson, "Beiträge zur Topologie der Gruppenmanniigfaltigkeiten," *ibid.*, vol. 42 (1941), pp. 1091-1137.

[14] H. C. Wang, "Homogeneous spaces with non-vanishing Euler characteristic," *ibid.*, vol. 50 (1949), pp. 925-953.

[15] ———, "Closed manifolds with homogeneous complex structure," *American Journal of Mathematics*, vol. 76 (1954), pp. 1-32.

[16] ———, "A new characterization of spheres of even dimension," *Koninklijke Nederlandse Akademie van Wetenschappen. Proceedings*, vol. 52 (1949), pp. 838-845.

[17] J. Poncet, "Groupes de Lie compacts de transformations de l'espaces euclidien et les sphères comme espaces homogènes," *Commentarii Mathematici Helvetici*, vol. 33 (1959), pp. 109-120.

[18] P. S. Mortert, "On a compact Lie group acting on a sphere" (abstract), *American Mathematical Society Notices*, vol. 5 (1958), p. 498.

[19] N. Nagano, "On some compact transformation groups on spheres," Scientific papers of College of General Education, University of Tokyo, vol. 9 (1959), pp. 213-218.

CONFORMAL TRANSFORMATIONS OF RIEMANN SURFACES.*

By Leon Greenberg.

1. **Introduction.** Let $\mathfrak{S}$ be a compact Riemann surface of genus $p > 1$, and let $\boldsymbol{C}(\mathfrak{S})$ be the group of 1-1 conformal transformations of $\mathfrak{S}$ onto itself. It is known that $\boldsymbol{C}(\mathfrak{S})$ is finite and that any finite group is isomorphic to a subgroup of some $\boldsymbol{C}(\mathfrak{S})$ (see [3]). The latter fact can be seen as follows. Let $\boldsymbol{G}$ be a finite group generated by $g_1, g_2, \cdots, g_n$, and let $\boldsymbol{F}$ be the group $\langle a_1, b_1, \cdots, a_n, b_n; \prod_{k=1}^{n} a_k b_k a_k^{-1} b_k^{-1} = 1\rangle$. There is a homomorphism $\phi: \boldsymbol{F} \to \boldsymbol{G}$, defined by $\phi(a_k) = g_k$, $\phi(b_k) = g_k^{-1}$. If $\boldsymbol{K}$ is the kernel, then $\boldsymbol{G}$ is isomorphic to $\boldsymbol{F}/\boldsymbol{K}$. $\boldsymbol{F}$ is the fundamental group of a compact Riemann surface $\mathfrak{T}$ of genus n. $\mathfrak{T}$ is covered by the half-plane $\mathfrak{H} = \{(x, y) \mid y > 0\}$, and the group of covering transformations is a discrete group of linear fractional transformations, isomorphic to $\boldsymbol{F}$. If we identify points of $\mathfrak{H}$ which are congruent under $\boldsymbol{K}$, we obtain a Riemann surface $\mathfrak{S} = \mathfrak{H}/\boldsymbol{K}$. $\mathfrak{S}$ is compact, since it is a finite covering space of $\mathfrak{T} = \mathfrak{H}/\boldsymbol{F}$ (the index $[\boldsymbol{F}: \boldsymbol{K}] = \text{order}(\boldsymbol{G})$ is finite). A conformal transformation c of $\mathfrak{S}$ can be lifted to a conformal transformation $\bar{c}$ of $\mathfrak{H}$. The transformation $\bar{c}$ maps an orbit $\boldsymbol{K}z$, $z \in \mathfrak{H}$, onto another orbit. From this it follows that $\bar{c}\boldsymbol{K} = \boldsymbol{K}\bar{c}$, or $\bar{c}$ is in the normalizer $\boldsymbol{N}[\boldsymbol{K}, \boldsymbol{C}(\mathfrak{H})]$ of $\boldsymbol{K}$ in $\boldsymbol{C}(\mathfrak{H})$. $\boldsymbol{C}(\mathfrak{S})$ is isomorphic to $\boldsymbol{N}[\boldsymbol{K}, \boldsymbol{C}(\mathfrak{H})]/\boldsymbol{K}$, which contains $\boldsymbol{F}/\boldsymbol{K}$ as a subgroup. Thus $\boldsymbol{G}$ is isomorphic to a subgroup of $\boldsymbol{C}(\mathfrak{S})$.

Since the fundamental groups of the non-compact surfaces are free groups, a similar argument shows that any finite group is realized as a subgroup of $\boldsymbol{C}(\mathfrak{S})$, where $\mathfrak{S}$ is some finite, non-compact Riemann surface (we shall call $\mathfrak{S}$ finite, if it has a finitely generated fundamental group). The purpose of this article is to show:

(a) If $\mathfrak{S}$ is a finite, non-compact Riemann surface which is not homeomorphic to a plane or cylinder, then $\boldsymbol{C}(\mathfrak{S})$ is finite.

(b) If $\boldsymbol{G}$ is a finite group, then there is a finite, non-compact Riemann surface $\mathfrak{S}$, such that $\boldsymbol{C}(\mathfrak{S})$ is isomorphic to $\boldsymbol{G}$. If $\boldsymbol{G}$ is a countable group, then there is a non-compact Riemann surface $\mathfrak{S}$, such that $\boldsymbol{C}(\mathfrak{S})$ is isomorphic to $\boldsymbol{G}$.

* Received September 4, 1959.

2. The hyperbolic plane. The half-plane $\mathfrak{H}$ can be given the Riemannian metric $ds^2 = (dx^2 + dy^2)/y^2$, so that it becomes the Poincaré model of the hyperbolic plane. Let X denote the x-axis, together with the point ∞. The geodesics in $\mathfrak{H}$ are semi-circles and half-lines orthogonal to X. The orientation-preserving isometries are exactly the conformal transformations of $\mathfrak{H}$. These are the linear fractional transformations

$$f(z) = (az + b)/(cz + d),$$

where $ad - bc = 1$ and a, b, c, d are real. We shall often write f as the matrix

$$f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

If f is hyperbolic, the geodesic A_f through the fixed points of f is invariant under f, and is called the axis of f. The translation length along A_f is

$$\lambda(f) = 2 \log \tfrac{1}{2}[t + (t^2 - 4)^{\frac{1}{2}}], \tag{1}$$

where $t = a + d$. In later calculations we shall be dealing with the case that $A_f = \{(x, y) \mid x^2 + y^2 = r^2, y > 0\}$. If $\lambda = \lambda(f)$, then

$$f = \begin{pmatrix} \cosh(\lambda/2) & r \sinh(\lambda/2) \\ (1/r)\sinh(\lambda/2) & \cosh(\lambda/2) \end{pmatrix}. \tag{2}$$

Corresponding to each isometry f, which moves the point ∞, there are a pair of (circular) geodesics I_f, I_f' which are the isometric circles of f. The transformation f is the product of a reflection in I_f and a reflection in the perpendicular bisector of the line segment between the centers of I_f and I_f'. The isometric circles of f do not intersect, if and only if f is hyperbolic. In this case I_f and I_f' are orthogonal to A_f. When $A_f = \{(x, y) \mid x^3 + y^2 = r^2, y > 0\}$, the isometric circles are of the forms:

$$I_f = \{(x, y) \mid (x + c)^2 + y^2 = s^2, y > 0\},$$
$$I_f' = \{(x, y) \mid (x - c)^2 + y^2 = s^2, y > 0\},$$

where $r^2 + s^2 = c^2$, since I_f and I_f' are orthogonal to A_f. Each of the pairs $\{A_f, \lambda(f)\}$, $\{A_f, I_f\}$ determines f completely. The correspondence between $\lambda = \lambda(f)$ and I_f is given by:

$$\begin{aligned} \lambda &= \log[(c + r)/(c - r)], \\ c &= r \coth(\lambda/2). \end{aligned} \tag{3}$$

(The radius s is determined by: $r^3 + s^2 = c^2$.)

Let I be a geodesic in $\mathfrak{H}$. For each point $z \in \mathfrak{H}$, there is a unique geodesic

I', which passes through z and is orthogonal to I. Let $\operatorname{proj}(z, I) = I \cap I'$. For any set $S \subset \mathfrak{H}$ let $\operatorname{proj}(S, I) = \{\operatorname{proj}(z, I) \mid z \in S\}$. If this latter set has a hyperbolic length, we shall denote this length by $\pi(S, I)$. Now let I and J be geodesics which intersect X in the points $\{i_1, i_2\}$ and $\{j_1, j_2\}$ respectively. It is not hard to show that

$$\pi(I, J) = \pi(J, I) = |\log |R(i_1, i_2; j_1, j_2)||, \tag{4}$$

where $R(i_1, i_2; j_1, j_2)$ is the cross ratio $[(i_1 - j_2)/(i_1 - j_1)][(i_2 - j_1)/(i_2 - j_2)]$. If I is contained in the interior of J or J in the interior of I (that is, if either $j_1 < i_1 < i_2 < j_2$ or $i_1 < j_1 < j_2 < i_2$), then (4) becomes

$$\pi(I, J) = \pi(J, I) = \log R(i_1, i_2; j_1, j_2). \tag{5}$$

3. **Discrete groups.** Let $\boldsymbol{F}$ be a subgroup of $\boldsymbol{C}(\mathfrak{H})$. The set of limit points of $\boldsymbol{F}$, denoted $L_{\boldsymbol{F}}$, is the intersection with X of the set of limit points of $\{f(z) \mid f \in \boldsymbol{F}\}$, where z is any point in $\mathfrak{H}$. $L_{\boldsymbol{F}}$ is independent of $z \in \mathfrak{H}$ and is invariant under $\boldsymbol{F}$. $L_{\boldsymbol{F}}$ can be the empty set, a single point, two points, X, or a perfect, nowhere dense subset of X. The latter occurs in the case that $\boldsymbol{F}$ is the group of covering transformations corresponding to a non-compact Riemann surface which is not homeomorphic to a plane or cylinder. In this case, $X - L_{\boldsymbol{F}}$ is composed of a countable number of open intervals, called regular intervals. Each regular interval is spanned by a geodesic, which may or may not be an axis of an element in $\boldsymbol{F}$. J. Nielsen [5], [6] has shown that if $\boldsymbol{F}$ is a finitely generated, discrete group, then each of these geodesics is an axis (a bounding axis) and there are a finite number of bounding axes $B_1, B_2, \cdots, B_m$, such that all bounding axes are contained in the set $\{fB_k \mid f \in \boldsymbol{F}, k = 1, 2, \cdots, m\}$.

A discrete group $\boldsymbol{F}$ has a canonical fundamental region, denoted $\mathfrak{R}_{\boldsymbol{F}}$, which consists of the region in $\mathfrak{H}$ outside of all isometric circles of elements in $\boldsymbol{F}$. The following is proved in [4, p. 76].

THEOREM 1. *Let $\boldsymbol{F}$ and $\boldsymbol{G}$ be discrete subgroups of $\boldsymbol{C}(\mathfrak{H})$ such that the isometric circles of $\boldsymbol{F}$ are contained in $\mathfrak{R}_{\boldsymbol{G}}$ and the isometric circles of $\boldsymbol{G}$ are contained in $\mathfrak{R}_{\boldsymbol{F}}$. Then the group $\boldsymbol{E}$ generated by $\boldsymbol{F}$ and $\boldsymbol{G}$ is discrete. $\boldsymbol{E}$ is the free product $\boldsymbol{F} * \boldsymbol{G}$ and $\mathfrak{R}_{\boldsymbol{E}} = \mathfrak{R}_{\boldsymbol{F}} \cap \mathfrak{R}_{\boldsymbol{G}}$.*

A method of constructing a class of discrete groups is the following. Let $I_1, I_1', I_2, I_2', \cdots, I_n, I_n'$ be (circular) geodesics which are external to each other and such that $\operatorname{radius}(I_k) = \operatorname{radius}(I_k')$. Let f_k be the linear fractional transformation with isometric circles I_k, I_k'. Let D_k (D_k') be the closed half-disk bounded by X and I_k (I_k'). Theorem 1 shows that the group $\boldsymbol{F}$ generated

by $\{f_k \mid k=1,2,\cdots,n\}$ is a discrete, free group of rank n, for which $\mathfrak{R}_F = \mathfrak{H} - \bigcup_{k=1}^{n} (D_k \cup D_k')$. It is not hard to show that the elements of F are all hyperbolic.

We shall need the following definitions. Let F be any group. An N-chain is a sequence of subgroups $F_1, F_2, \cdots, F_n$ such that

(a) $F_k \neq \{1\}$, $k=1,2,\cdots,n$,

(b) Either F_k is a normal subgroup of F_{k+1} or F_{k+1} is a normal subgroup of F_k, $k=1,2,\cdots,n-1$.

Two subgroups G and H are N-equivalent if there is an N-chain $G=F_1, F_2, \cdots, F_n = H$. G is an N-subgroup if it is N-equivalent to F. G is N-maximal if it contains every subgroup of F which is N-equivalent to G. The following has been proved in [2, Theorem 4].

THEOREM 2. *Let F be a discrete subgroup of $C(\mathfrak{H})$ and let G be a finitely generated N-subgroup of F. Then the index $[F:G]$ is finite.*

LEMMA 1. *Let F be a discrete, non-commutative subgroup of $C(\mathfrak{H})$. Then the normalizer $N[F, C(\mathfrak{H})]$ is discrete.*

Proof. Suppose that there is a sequence $\{n_k\} \subset N[F, C(\mathfrak{H})]$, such that $\lim_{k\to\infty} n_k = 1$. Then $\lim_{k\to\infty} f n_k f^{-1} n_k^{-1} = 1$, where f is any element of F. Since $f n_k f^{-1} n_k^{-1} \in F$, and F is discrete, it follows that $f n_k f^{-1} n_k^{-1} = 1$, for almost all k. Two linear fractional transformations commute, if and only if they have the same fixed points. This shows that F must be a commutative group, contrary to assumption.

THEOREM 3. *Let $\mathfrak{S}$ be a finite Riemann surface which is not homeomorphic to a sphere, plane, cylinder or torus. Then $C(\mathfrak{S})$ is finite.*

Proof. The restrictions on $\mathfrak{S}$ imply that $\mathfrak{S}$ is (conformally) covered by $\mathfrak{H}$. Let G be the group covering transformations and let F be the normalizer $N[G, C(\mathfrak{H})]$. $C(\mathfrak{S})$ is isomorphic to F/G. Since $\mathfrak{S}$ is not homeomorphic to a plane or cylinder, G is non-trivial and non-commutative. Lemma 1 implies that F is discrete, while Theorem 2 implies that the index $[F:G]$ is finite. Therefore $C(\mathfrak{S})$ is finite.

We now wish to show that for any finite group G, there is a finite Riemann surface $\mathfrak{S}$ so that $C(\mathfrak{S})$ is isomorphic to G. In order to prove this, we shall show that for any positive integer $n > 1$, there is an N-maximal, discrete (hyperbolic) group F_n which is a free group of rank n. Suppose, for the

moment, that we know such groups exist. If G is generated by n generators, then G is isomorphic to a factor group F_n/K. K is finitely generated, since the index $[F_n : K]$ is finite; therefore the Riemann surface $\mathfrak{S} = \mathfrak{H}/K$ is finite. $C(\mathfrak{S})$ is isomorphic to $N[K, C(\mathfrak{H})]/K$. Now F_n is N-equivalent to $N[K, C(\mathfrak{H})]$. Since F_n is N-maximal, it follows that $F_n = N[K, C(\mathfrak{H})]$, and $C(\mathfrak{S}) \approx F_n/K \approx G$.

Similarly, if we prove the existence of an N-maximal, discrete group, which is free of countably infinite rank, then we can show that any countable group is realized as the conformal group of some Riemann surface.

4. Construction of the group F_n. We shall now construct an N-maximal, discrete group F_n which is a free group of rank n. F_n will be generated by elements $f_1, \cdots, f_n$ with isometric circles $I_1, I_1', \cdots, I_n, I_n'$ and axes $A_1, \cdots, A_n$. The bounding axes of F_n will fall into a finite number of disjoint classes $F_nB_1, F_nB_2, \cdots, F_nB_n$. The minimal translation length λ along an axis A is the same for all axes in the class F_nA (since $gA_f = A_{gfg^{-1}}$ and $\lambda(f) = \lambda(gfg^{-1})$). Let λ_k be the minimal translation length for the class F_nB_k.

We shall first show that F_n can be constructed so that:

(α) The isometric circles $\{I_k, I_k' \mid k = 1, 2, \cdots, n\}$ are external to each other.

(β) λ_i/λ_j is irrational, if $i \neq j$.

(γ) $\lambda_1 < 2\pi(B_1, B_m)$.

In Section 6 we shall show that the resulting group F_n is N-maximal.

Let us first choose the axes $\{A_k \mid k = 1, 2, \cdots, n\}$, and then choose I_k, I_k' as orthocircles with equal radii, and orthogonal to A_k. We shall choose $A_k = \{(x, y) \mid x^2 + y^2 = r_k^2, y > 0\}$, where $r_1 = 2$, $r_n = 1$ and $\{r_k \mid k = 2, 3, \cdots, n-1\}$ are to be determined, and will satisfy

$$2 = r_1 > r_2 > \cdots > r_n = 1. \tag{6}$$

The isometric circles will be:

$$I_k = \{(x, y) \mid (x + c_k)^2 + y^2 = s_k^2, y > 0\}$$

and

$$I_k' = \{(x, y) \mid (x - c_k)^2 + y^2 = s_k^2, y > 0\},$$

where the $\{c_k\}$ and $\{s_k\}$ are to be determined and will satisfy

$$r_k^2 + s_k^2 = c_k^2; \tag{7}$$

in order that I_k and I_k' be orthogonal to A_k, and

$$(8)\qquad c_1 - s_1 > c_2 + s_2 > c_2 - s_2 > c_3 + s_3 > \cdots > c_n - s_n > 0,$$

in order that the isometric circles be external to each other. For any choice of $\{r_k, c_k, s_k\}$ satisfying (6), (7), (8), the surface $\mathfrak{H}/F_n$ is homeomorphic to a sphere from which $n+1$ closed disks have beeen removed. This means that there are exactly $n+1$ congruence classes of bounding axes, or $m = n+1$. The fundamental region $\mathfrak{R}_{F_n}$ intersects X in the intervals:

$$\begin{aligned} W_1^{(0)} &= X - (-c_1 - s_1, c_1 + s_1), \\ V_{-k}^{(0)} &= [-c_{k-1} + s_{k-1}, -c_k - s_k], \quad k = 2, 3, \cdots, n, \\ V_k^{(0)} &= [c_k + s_k, c_{k-1} - s_{k-1}], \quad k = 2, 3, \cdots, n, \\ W_{n+1}^{(0)} &= [-c_n + s_n, c_n - s_n]. \end{aligned}$$

Each of these intervals is contained in a regular interval, and any regular interval is congruent to a regular interval which contains at least one of these intervals.

Let $W_1^{(h)} = f_1^h W_1^{(0)}$ $(h = 0, \pm 1, \pm 2, \cdots)$. The interval $W_1 = \bigcup_{h=-\infty}^{\infty} W_1^{(h)}$ is bounded by the limit points $\{-r_1, r_1\}$. Thus $W_1^{(0)}$ is contained in the regular interval W_1, which is spanned by the bounding axis A_1, and we can take $B_1 = A_1$. Similarly, if $W_{n+1}^{(h)} = f_n^h W_{n+1}^{(0)}$, then $W_{n+1} = \bigcup_{h=-\infty}^{\infty} W_{n+1}^{(h)}$ is a regular interval containing $W_{n+1}^{(0)}$, bounded by the limit points $\{-r_n, r_n\}$ and spanned by the bounding axis A_n. We can take $B_{n+1} = A_n$.

Now $f_{k-1}^{-1} V_k^{(0)}$, $V_{-k}^{(0)}$ and $f_k^{-1} V_k^{(0)}$ are three consecutive intervals with common endpoints, and $f_k^{-1} f_{k-1}$ maps the first interval into the third. Therefore if $W_k^{(0)} = V_{-k}^{(0)} \cup f_k^{-1} V_k^{(0)}$ and $W_k^{(h)} = (f_k^{-1} f_{k-1})^h W_k^{(0)}$, then $W_k = \bigcup_{h=-\infty}^{\infty} W_k^{(h)}$ is a regular interval which is spanned by the bounding axis $B_k = A_{f_k^{-1} f_{k-1}}$. Thus we have found representatives for the classes of bounding axes:

$$B_1 = A_1, B_2 = A_{f_2^{-1} f_1}, \cdots, B_n = A_{f_n^{-1} f_{n-1}}, B_{n+1} = A_n.$$

We are now ready to begin choosing the f_k so that conditions (α), (β) and (γ) are satisfied. We first choose f_1 so that $\lambda(f_1) < 2\pi(A_1, A_n)$ and the points -1 and 1 are outside the isometric circles I_1, I_1'. This latter condition is exactly $\lambda(f_1) > \pi(A_1, A_n)$. Since $\pi(A_1, A_n) = \log R(-1, 1; -2, 2) = \log 9$, the condition on f_1 is

$$\log 9 < \lambda(f_1) < 2 \log 9.$$

Choose any number λ between $\log 9$ and $2\log 9$ and define

$$f_1 = \begin{pmatrix} \cosh(\lambda/2) & 2\sinh(\lambda/2) \\ \frac{1}{2}\sinh(\lambda/2) & \cosh(\lambda/2) \end{pmatrix}.$$

Then $\lambda = \lambda(f_1) = \lambda_1$ satisfies condition (γ). Furthermore the condition $\lambda(f_1) > \pi(A_1, A_n)$ guarantees that $c_1 - s_1 > 1$.

Now suppose that we have found $f_1, f_2, \cdots, f_k$ $(k \leqq n-2)$ such that

$$c_1 - s_1 > c_2 + s_2 > c_2 - s_2 > \cdots > c_k - s_k > 1, \qquad 2 > r_2 > r_3 > \cdots > 1,$$

and such that the ratio of any two of the numbers

$$\lambda(f_1), \lambda(f_2^{-1}f_1), \cdots, \lambda(f_k^{-1}f_{k-1})$$

is irrational. We wish to extend these conditions to a transformation f_{k+1}. The condition $c_k - s_k > c_{k+1} + s_{k+1}$ is equivalent to

$$c_{k+1} < [(c_k - s_k)^2 + r_{k+1}^2]/2(c_k - s_k).$$

The condition $c_{k+1} - s_{k+1} > 1$ is equivalent to

$$c_{k+1} < (1 + r_{k+1}^2)/2.$$

Choose r_{k+1} so that $1 < r_{k+1} < (c_k - s_k)^{\frac{1}{2}}$. Then it follows that

$$(1 + r_{k+1}^2)/2 < [(c_k - s_k)^2 + r_{k+1}^2]/2(c_k - s_k) \text{ and } r_{k+1} < r_k.$$

Now if we choose any c_{k+1} so that

$$r_{k+1} < c_{k+1} < (1 + r_{k+1}^2)/2, \tag{9}$$

and define $s_{k+1} = (c_{k+1}^2 - r_{k+1}^2)^{\frac{1}{2}}$, it will follow that

$$c_k - s_k > c_{k+1} + s_{k+1} > c_{k+1} - s_{k+1} > 1.$$

However we wish to choose c_{k+1} so that the ratio of any two of the numbers $\lambda(f_1), \lambda(f_2^{-1}f_1), \cdots, \lambda(f_{k+1}^{-1}f_k)$ is irrational. The relations (3) show that (9) is equivalent to

$$2\log[(r_{k+1} + 1)/(r_{k+1} - 1)] < \lambda(f_{k+1}) < \infty. \tag{10}$$

Let $f(x) = \begin{pmatrix} \cosh(x/2) & r_{k+1}\sinh(x/2) \\ (1/r_{k+1})\sinh(x/2) & \cosh(x/2) \end{pmatrix}$ and $\psi(x) = \lambda(f(x)^{-1}f_k)$.

The function ψ is continuous and maps the interval

$$I = (2\log[(r_{k+1} + 1)/(r_{k-1} - 1)], \infty)$$

onto an interval J. Let

$$K_1 = \{r\lambda(f_1) \mid r \text{ rational}\}, \qquad K_i = \{r\lambda(f_i^{-1}f_{i-1}) \mid r \text{ rational}\},$$
$$i = 2, 3, \cdots, k.$$

The set $K = \bigcup_{i=1}^{k} K_i$ is countable, so $J - K$ is non-empty. Choose $x_0 \in I$ so that $\psi(x_0) \in J - K$, and define $f_{k+1} = f(x_0)$. The transformations $f_1, f_2, \cdots, f_{k+1}$ satisfy the conditions:

$$c_1 - s_1 > c_2 + s_2 > c_2 - s_2 > \cdots > c_{k+1} - s_{k+1} > 1,$$

$$2 = r_1 > r_2 > \cdots > r_{k+1} > 1,$$

and the ratio of any two of the numbers $\lambda(f_1), \lambda(f_2^{-1}f_1), \cdots, \lambda(f_{k+1}^{-1}f_k)$ is irrational.

Now suppose we have found $f_1, f_2, \cdots, f_{n-1}$ with the required properties. We wish to find f_n so that $r_n = 1$, $c_{n-1} - s_{n-1} > c_n + s_n > c_n - s_n > 0$ and condition (β) are fulfilled. The second condition is equivalent to

$$1 < c_n < u = [(c_{n-1} - s_{n-1})^2 + 1]/2(c_{n-1} - s_{n-1}),$$

which, as before, is equivalent to

$$\log[(u+1)/(u-1)] < \lambda(f_n) < \infty.$$

Let
$$f(x) = \begin{pmatrix} \cosh(x/2) & \sinh(x/2) \\ \sinh(x/2) & \cosh(x/2) \end{pmatrix}$$

and
$$\psi(x) = \lambda(f(x)^{-1} f_{n-1})/x.$$

The function ψ is continuous on the interval $I = (\log[(u+1)/(u-1)], \infty)$ and maps I onto an interval J. Define K_i as before and $K = \bigcup_{i=1}^{n-1} K_i$. Now the set $M = \{x \mid x \in I \text{ and } \psi(x) \text{ is irrational}\}$ is an uncountable set, while K is countable. Therefore M cannot be a subset of K, and $M \cap (I - K)$ is non-empty. Choose $x_0 \in M \cap (I - K)$ and let $f_n = f(x_0)$. Then the transformations $\{f_k \mid k = 1, 2, \cdots, n\}$ satisfy conditions (α), (β), (γ).

5. Construction of the group F_∞. In this section we shall construct a discrete group F_∞ which is a free group of rank $\aleph_0$, and which will later be shown to be N-maximal. The construction will be similar to that in the previous section. F_∞ will be generated by transformations $\{f_k \mid k = 0, 1, 2, \cdots\}$, with isometric circles $\{I_k, I_k' \mid k = 0, 1, 2, \cdots\}$ and axes $\{A_k \mid k = 0, 1, 2, \cdots\}$. We can choose $A_k = \{(x, y) \mid x^2 + y^2 = r_k^2, y > 0\}$, where $r_0 = \frac{1}{2}$, $r_1 = 2$, and $r_1 > r_2 > r_3 > \cdots > 1$. The isometric circles will be

$$I_k = \{(x, y) \mid (x + c_k)^2 + y^2 = s_k^2, y > 0\},$$

$$I_k' = \{(x, y) \mid (x - c_k)^2 + y^2 = s_k^2, y > 0\},$$

where $r_k^2 + s_k^2 = c_k^2$,

$$c_1 - s_1 > c_2 + s_2 > c_2 - s_2 > \cdots > c_k + s_k > c_k - s_k > \cdots$$
$$> 1 > c_0 + s_0 > c_0 - s_0 > 0,$$

and $\lim_{k\to\infty} c_k = 1$.

For any choice of $\{r_k, c_k, s_k\}$, satisfying the above conditions, the fundamental region $\mathfrak{R}_{F_\infty}$ intersects X in the intervals:

$$W_0^{(0)} = [-c_0 + s_0, c_0 - s_0],$$
$$W_1^{(0)} = X - (-c_1 - s_1, c_1 + s_1),$$
$$W_- = [-1, -c_0 - s_0],$$
$$W_+ = [c_0 + s_0, 1],$$
$$V_{-k}^{(0)} = [-c_{k-1} + s_{k-1}, -c_k - s_k], \quad (k = 2, 3, \cdots),$$
$$V_k^{(0)} = [c_k + s_k, c_{k-1} - s_{k-1}], \quad (k = 2, 3, \cdots).$$

Any regular interval is congruent, under F_∞, to an interval which contains at least one of the above intervals.

Let $W_0^{(h)} = f_0^h W_0^{(0)}$, $W_0 = \bigcup_{h=-\infty}^{\infty} W_0^{(h)}$, $W_1^{(h)} = f_1^h W_1^{(0)}$ and $W_1 = \bigcup_{h=-\infty}^{\infty} W_1^{(h)}$. Then W_0 and W_1 are regular intervals, spanned by the bounding axes $B_0 = A_0$, $B_1 = A_1$, respectively. Let $W_k^{(0)} = V_{-k}^{(0)} \cup f_k^{-1} V_k^{(0)}$, $W_k^{(h)} = (f_k^{-1} f_{k-1})^h W_k^{(0)}$ and $W_k = \bigcup_{h=-\infty}^{\infty} W_k^{(h)}$. Then W_k is a regular interval, spanned by the bounding axis $B_k = A_{f_k^{-1} f_{k-1}}$. The interval $W = W_- \cup f_0^{-1} W_+$ is a regular interval, bounded by the limit points $\{-1, f_0^{-1}(1)\}$. The geodesic which spans W is not an axis of F_∞. Every regular interval is contained in the set $\{fW, fW_k \mid f \in F_\infty, k = 0, 1, 2, \cdots\}$, and every bounding axis is contained in $\{fB_k \mid f \in F_\infty, k = 0, 1, 2, \cdots\}$. Let $\lambda_0 = \lambda(f_0)$, $\lambda_1 = \lambda(f_1)$ and $\lambda_k = \lambda(f_k^{-1} f_{k-1})$ $(k = 2, 3, \cdots)$. In the same way as in the case of $\boldsymbol{F}_n$, we can show that $\{f_k \mid k = 0, 1, 2, \cdots\}$ can be chosen so that:

(α) The isometric circles $\{I_k, I_k' \mid k = 0, 1, 2, \cdots\}$ are external to each other.

(β) λ_i/λ_j is irrational, if $i \neq j$.

(γ) $\lambda_1 < 2\pi(B_1, B_0)$.

6. The N-maximality of F_n and F_∞. The following is proved in [1, p. 43].

LEMMA 2. *Let G be a discrete subgroup of $C(\mathfrak{H})$. If Y is a closed subset of X, which contains more than one point and is invariant under G, then $Y \supset L_G$.*

LEMMA 3. *Let E and F be subgroups of $C(\mathfrak{H})$. Suppose that F is discrete, L_F contains more than two points*[1] *and E is N-equivalent to F. Then $L_E = L_F$.*

Proof. There is an N-chain $F = G_1, G_2, \cdots, G_m = E$. We shall prove inductively that G_k is discrete and $L_{G_k} = L_F$.

Let us assume the above is true for G_k, and prove it for G_{k+1}. Either G_k is a normal subgroup of G_{k+1} or G_{k+1} is a normal subgroup of G_k. We shall first assume the former. Then Lemma 1 implies that G_{k+1} is discrete. Since $G_{k+1} \supset G_k$, $L_F = L_{G_k} \subset L_{G_{k+1}}$. We need to show that $L_{G_{k+1}} \subset L_{G_k}$. To do this, we shall show that L_{G_k} is invariant under G_{k+1}.

If $z_0 \in L_{G_k}$, there is a sequence $\{g_m\} \subset G_k$, so that $\lim_{m\to\infty} g_m(z) = z_0$, for any point $z \in \mathfrak{H}$. Let $g \in G_{k+1}$. Then $\lim_{m\to\infty} g_m g^{-1}(z) = z_0$ and $\lim_{m\to\infty} g g_m g^{-1}(z) = g(z_0)$, since g is continuous. But $g g_m g^{-1} \in G_k$, so $g(z_0) \in L_{G_k}$, and L_{G_k} is invariant under G_{k+1}. Furthermore $L_{G_k} = L_F$ contains more than one point. Lemma 2 (with $Y = L_{G_k}$ and $G = G_{k+1}$) now implies that $L_{G_k} \subset L_{G_{k+1}}$. Therefore $L_{G_{k+1}} = L_{G_k} = L_F$.

Now suppose that G_{k+1} is a normal subgroup of G_k. It is immediate that G_{k+1} is discrete and $L_{G_{k+1}} \subset L_{G_k}$. The argument given above (reversing the roles of G_k and G_{k+1}) shows that $L_{G_{k+1}}$ is invariant under G_k. If $L_{G_{k+1}}$ is the empty set or $L_{G_{k+1}}$ consists of one point, then G_{k+1} is an elliptic or parabolic group. In this case there is a point $z \in \mathfrak{H}$ or X, which is the common fixed point (in $\mathfrak{H}$ or X) of all elements in G_{k+1}. If $g \in G_k$ and $h \in G_{k+1}$, then $g(z)$ is a fixed point of $ghg^{-1} \in G_{k+1}$. It follows that z must be a common fixed point for all elements of G_k. Then G_k is an elliptic or parabolic group. This contradicts the assumption that L_{G_k} $(= L_F)$ contains more than two points. Therefore $L_{G_{k+1}}$ contains more than one point, and Lemma 2 implies that $L_{G_{k+1}} \supset L_{G_{k+1}}$. It follows that $L_{G_{k+1}} = L_{G_k}$ and $L_E = L_F$.

Let $G_n = \{g \mid g \in C(\mathfrak{H})$ and $gL_{F_n} = L_{F_n}\}$ and $G_\infty = \{g \mid g \in C(\mathfrak{H})$ and $gL_{F_\infty} = L_{F_\infty}\}$.

LEMMA 4. *G_n and G_∞ are discrete.*

Proof. A closed subgroup of $C(\mathfrak{H})$ must be one of the following: 1) the group of all hyperbolic transformations with a given pair of fixed points in

[1] The Lemma is true without this assumption, but we are interested only in this case.

X; 2) the group of all parabolic transformations with a given fixed point in X; 3) the group of all elliptic transformations with a given fixed point in $\mathfrak{H}$; 4) the group of all hyperbolic and parabolic transformations with a given fixed point in X; 5) the group of all hyperbolic and elliptic transformations which leave a given geodesic in $\mathfrak{H}$ invariant; 6) $\boldsymbol{C}(\mathfrak{H})$; 7) a discrete group.

The only possibility for the closures $\bar{\boldsymbol{G}}_n$ and $\bar{\boldsymbol{G}}_\infty$ is 7). Therefore $\boldsymbol{G}_n$ and $\boldsymbol{G}_\infty$ are discrete.

LEMMA 5. *$\boldsymbol{G}_n = \boldsymbol{F}_n$ and $\boldsymbol{G}_\infty = \boldsymbol{F}_\infty$.*

Proof. It is clear that $\boldsymbol{F}_n \subset \boldsymbol{G}_n$ and $\boldsymbol{F}_\infty \subset \boldsymbol{G}_\infty$, since L_F is invariant under $\boldsymbol{F}$, for any group $\boldsymbol{F} \subset \boldsymbol{C}(\mathfrak{H})$. We must prove that $\boldsymbol{F}_n \supset \boldsymbol{G}_n$ and $\boldsymbol{F}_\infty \supset \boldsymbol{G}_\infty$. We shall first consider the case of $\boldsymbol{F}_n$.

Let $g \in \boldsymbol{G}_n$. The regular interval W_1, spanned by B_1, is mapped by g onto another regular interval Y, spanned by an axis $A_{f'} = gA_1$, where $f' \in \boldsymbol{F}_n$. Now Y is in one of the classes $\{\boldsymbol{F}_n W_k | k = 1, 2, \cdots, n+1\}$. Suppose Y is in $\boldsymbol{F}_n W_k$, so that $gW_1 = fW_k$ and $gB_1 = fB_k$, where $f \in \boldsymbol{F}_n$. Let g_1, g_k be primary elements of $\boldsymbol{G}_n$ with axes B_1, B_k respectively (such elements exist since $\boldsymbol{G}_n$ is discrete). Then since B_1 and B_k are congruent in $\boldsymbol{G}$, it follows that $\lambda(g_1) = \lambda(g_k)$. Furthermore $\lambda_1 = \lambda(f_1)$ must be an integral multiple of $\lambda(g_1)$, say $\lambda_1 = p\lambda(g_1)$. Similarly $\lambda_k = \lambda(f_k^{-1}f_{k-1}) = q\lambda(g_k)$. Therefore $\lambda_1/\lambda_k = p/q$, which is contrary to the construction of $\boldsymbol{F}_n$, unless $k = 1$. Thus, we have found that $gB_1 = fB_1$ (and $gW_1 = fW_1$). The transformation $f^{-1}g$ maps W_1 onto itself, and is therefore a hyperbolic transformation with axis B_1. Therefore $f^{-1}g = g_1^k$, for some integer k, or $g = fg_1^k$. This implies that the index $[\boldsymbol{G}_n : \boldsymbol{F}_n]$ is finite, so $\boldsymbol{G}_n$ is N-equivalent to $\boldsymbol{F}_n$. Therefore, according to Lemma 3, $L_{F_n} = L_{G_n}$.

We now note that the condition $\lambda_1 < 2\pi(B_1, B_{n+1})$ implies that $\lambda(g_1) = \lambda(f_1)$, or $g_1 = f_1^{\pm 1}$. For suppose g is a transformation with axis B_1 such that $\lambda(g) < \pi(B_1, B_{n+1})$. Then B_{n+1} intersects $g(B_{n+1})$, so that $gf_{n+1}g^{-1}$ has a fixed point (which is a limit point) in the regular interval W_{n+1}. Since W_{n+1} is a regular interval for $\boldsymbol{G}_n$ (for $L_{F_n} = L_{G_n}$), no such element g can belong to $\boldsymbol{G}_n$. Now $\lambda(f_1) = p\lambda(g_1)$, for some integer p. If $p > 1$, then $\lambda(g_1) = \lambda(f_1)/p \leqq \lambda(f_1)/2 < \pi(B_1, B_{n+1})$. We conclude that $g_1 = f_1^{\pm 1}$. Since for any $g \in \boldsymbol{G}_n$ there is $f \in \boldsymbol{F}_n$ and an integer k so that $g = fg_1^k$, it now follows that $\boldsymbol{G}_n = \boldsymbol{F}_n$.

The argument for $\boldsymbol{F}_\infty$ is similar. The only difference is that gW_1 might be fW rather than fW_k. This can happen only if the geodesic, spanning W, is an axis of an element in $\boldsymbol{G}_\infty$ (even though it is not an axis of an element in $\boldsymbol{F}_\infty$). We shall show that this cannot happen, by showing that the index

$r = [G_\infty : F_\infty]$ is finite. If this is true, and $g \in G_\infty$, then one of the elements $g, g^2, \cdots, g^r$ must be in F_∞. These elements all have the same axis, so we see that for any $g \in G_\infty$ there is an element in F_∞ whose axis coincides with A_g.

If W_1 is not congruent to W under G_∞, then the argument used for F_n is valid here. Suppose then, that W_1 is congruent to W. W_2 cannot be congruent to W under G_∞. For in this case W_2 would be congruent to W_1, and λ_2/λ_1 would be rational.

An element $g \in G_\infty$ maps W_2 onto a regular interval W', which is in some class $F_\infty W_k$. We now have $gW_2 = fW_k$, where $f \in F_\infty$. By the same argument as before, we see that $k = 2$, because λ_2/λ_k is irrational. Therefore $gW_2 = fW_2$ and $f^{-1}g$ maps W_2 onto itself. If g_2 is a primary element in G_∞, with axis B_2, then $f^{-1}g = g_2{}^m$. If $\lambda_2 = n\lambda(g_2)$, then every element in G_∞ is congruent mod F_∞ to one of the elements $1, g_2, \cdots, g_2{}^{n-1}$. Thus the index $[G_\infty : F_\infty]$ is finite, and as we have seen, this implies that $G_\infty = F_\infty$.

THEOREM 4. *The groups F_n and F_∞ are N-maximal.*

Proof. Let F be one of the groups F_n or F_∞. If E is N-equivalent to F, Lemma 3 implies that $L_E = L_F$. Therefore L_F is invariant under E, and $E \subset G = \{g \mid g \in C(\mathfrak{H}) \text{ and } gL_F = L_F\}$. But according to Lemma 5, $G = F$. Therefore $E \subset F$, and F is N-maximal.

The previous theorem and the remarks at the end of Section 3 now imply the following.

THEOREM 5. *Let G be a countable group. Then there is a Riemann surface $\mathfrak{S}$, such that $C(\mathfrak{S})$ is isomorphic to G. If G is finite, the surface $\mathfrak{S}$ can be taken to be finite.*

BROWN UNIVERSITY.

REFERENCES.

[1] L. Ford, *Automorphic Functions*, New York, 1951.

[2] L. Greenberg, "Discrete groups of motions," *Canadian Journal of Mathematics*, vol. 12 (1960), pp. 415-426.

[3] A. Hurwitz, "Ueber algebraische Gebilde mit eindeutigen Transformationen in sich," *Mathematische Annalen*, Band 41 (1893), pp. 403-442.

[4] W. Magnus, *Discrete groups*, New York University notes, 1952.

[5] J. Nielsen, "Über Gruppen linearer Transformationen," *Mittelungen der Mathematischen Gesellschaft in Hamburg*, Band VIII (1940), pp. 82-104.

[6] ———, "Nogle grundlaegende begreber vedrørende diskontinuerte grupper af lineaere substitutioner i en kompleks variabel," *Den 11'e Skandinaviske Matematikerkongress i Trondheim* (1949), pp. 61-70.

SUBGROUPS OF THE MODULAR GROUP AND SUMS OF SQUARES.* [1]

By Morris Newman.

Introduction. The problem of counting the number of representations $r_s(n)$ of a positive integer n as the sum of a fixed number of squares s has been of interest to mathematicians for many years. This function is defined as follows: If n is a non-negative integer, then $r_s(n)$ is the coefficient of q^n in

$$\{\sum_{n=-\infty}^{\infty} q^{n^2}\}^s;$$

otherwise, $r_s(n)$ is zero. The results obtained have generally been of two kinds: Asymptotic formulas, showing the behaviour of $r_s(n)$ for large n and fixed s; and exact formulas, which express $r_s(n)$ in terms of other arithmetical functions such as divisor sums and which are considered to be more elementary than $r_s(n)$. There is yet a third kind of relationship which has occurred occasionally in the literature but has not been fully exploited. This is the recurrence formula, an instance of which is

$$r_3(np^2) - r_3(n) = (p - (-n/p))(r_3(n) - r_3(n/p^2)),$$

where p is an odd prime, n an arbitrary integer and $(-n/p)$ the usual quadratic reciprocity symbol. Such formulas give information about the internal structure of the numbers $r_s(n)$.

The principal purpose of this paper is to prove that recurrence formulas of this type exist for all positive integers s and all odd primes p. The number of terms in these formulas depends only on s and p. The method used is the 'subgroup' method which concerns itself with functions invariant with respect to the substitutions of a suitable subgroup of the modular group Γ, regarded as linear fractional transformations. Related problems have been treated by the author in [6] and [10] by the same method.

The paper falls naturally into three parts. In § 1 the necessary group-

* Received October 9, 1959.

[1] The preparation of this paper was supported (in part) by the Office of Naval Research.

theoretic preliminaries are developed. These are given in much greater detail than is actually required for the remainder of the paper. In addition several theorems relevant to the structure of modular subgroups are proved which are not used in the remainder of the paper. These theorems are of high intrinsic interest however and generalize previous theorems given in [8] and [9]. In § 2 the necessary transformation equations for the associated modular forms are derived. In § 3 the class of modular functions to be studied is defined and the properties of these functions worked out, leading to a proof of the theorem mentioned above. Finally, it is shown how a Hauptmodul for a certain subgroup of Γ can be constructed from these functions.

1. Subgroups of the modular group. In what follows Γ is the full modular group; that is, the totality of rational integral 2×2 matrices of determinant 1. Γ is generated by the matrices

$$T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

with defining relations

$$T^2 = (ST)^3 = -I,$$

where I is the identity matrix. We also define

$$W = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Then for all integral k,

$$S^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \qquad W^k = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}.$$

Furthermore

$$W^{-1}T = TS, \qquad W = STS.$$

The group generated by the elements A, B of Γ will be denoted by $[A, B]$. If A is the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and τ is a complex number, then by $A\tau$ we shall understand

$$(a\tau + b)/(c\tau + d).$$

We shall say that $g(\tau)$ is a function on a subgroup G of Γ, if $g(A\tau) = g(\tau)$ for all $A \in G$.

Let G be a subgroup of Γ. We define a class of congruence subgroups $G(n)$, n a positive integer, as follows: The element

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

of G belongs to $G(n)$ if and only if $c \equiv 0 \pmod{n}$. If $G = \Gamma$, then the groups so defined become the customary congruence subgroups $\Gamma_0(n)$. We shall assume that G does not consists solely of powers of S, since then $G(n) = G$ and the groups are without interest. If m and n are positive integers and $[m, n]$ is their least common multiple, then

$$G(m) \cap G(n) = G([m, n]). \tag{1.1}$$

Formula (1.1) implies that if $m \mid n$, then $G(m) \supset G(n)$. The converse is not necessarily true. We can say however that if $G(m) \supset G(n)$, then $m \mid n\delta$, where δ is the greatest common divisor of all integers c such that

$$\begin{bmatrix} a & b \\ nc & d \end{bmatrix} \in G(n).$$

For example, suppose that $m > 1$, $(m, n) = 1$ and take $G = \Gamma_0(m)$. Then $G(m) = \Gamma_0(m)$, $G(n) = \Gamma_0(mn)$. Thus $G(m) \supset G(n)$ but m does not divide n. Here $\delta = m$ and certainly $m \mid n\delta$.

We now define the integers α, β as follows: If some non-zero power of S is in G, let α be the least positive integer such that $S^\alpha \in G$; otherwise let $\alpha = 0$. If some nan-zero power of W is in G, let β be the least positive integer such that $W^\beta \in G$; otherwise let $\beta = 0$. We say that G is of type (α, β). For example, $\Gamma_0(n)$ is of type $(1, n)$, $\Gamma^0(n) = T^{-1}\Gamma_0(n)T$ is of type $(n, 1)$, $\Gamma_0{}^0(n) = \Gamma_0(n) \cap \Gamma^0(n)$ is of type (n, n) and the principal congruence subgroup of level n (which we denote by $\Gamma[n]$) is of type (n, n).

Every group of type (α, β) contains as a subgroup the group

$$G_{\alpha,\beta} = [S^\alpha, W^\beta];$$

and it is of interest to investigate these groups further. We have

THEOREM 1.1. *Suppose that α, β are positive. Then if $\alpha\beta > 4$, $G_{\alpha,\beta}$ is of infinite index in Γ. If $\alpha\beta \leqq 4$, $G_{\alpha,\beta}$ is of finite index in Γ and $G_{1,1} = \Gamma$, $G_{1,2} = \Gamma_0(2)$, $G_{1,3} = \Gamma_0(3)$, $G_{1,4} = \Gamma_0(4)$, $G_{2,1} = \Gamma^0(2)$, $G_{2,2} = \Gamma[2]$, $G_{3,1} = \Gamma^0(3)$, $G_{4,1} = \Gamma^0(4)$. Furthermore $G_{\alpha,\beta}$ is a free group for $\alpha\beta \geqq 4$.*

Proof. If $\alpha\beta > 4$, the fundamental region of $G_{\alpha,\beta}$ contains a segment of the real axis, which proves that $G_{\alpha,\beta}$ is of infinite index in Γ. Also it has

been shown in [1] and [3] that for $\alpha\beta \geqq 4$, $G_{\alpha,\beta}$ is a free group. We now suppose $\alpha\beta \leqq 4$. Because of the relationship

$$T^{-1}G_{\alpha,\beta}T = G_{\beta,\alpha}$$

we need only consider $\alpha \leqq \beta$. The group $G_{1,1}$ is clearly Γ, since $T = S^{-1}WS^{-1}$. Suppose $2 \leqq n \leqq 4$. We have that $\Gamma_0(n) \supset G_{1,n}$. Let

$$A = \begin{bmatrix} a & b \\ nc & d \end{bmatrix} \in \Gamma_0(n).$$

Then

$$W^{nx}A = \begin{bmatrix} a & * \\ nc_0 & * \end{bmatrix}, \qquad c_0 = ax + c.$$

Since a cannot vanish, x can be determined so that $|c_0| \leqq \frac{1}{2}|a|$. Furthermore

$$S^{y}W^{nx}A = \begin{bmatrix} a_0 & * \\ nc_0 & * \end{bmatrix}, \qquad a_0 = a + nc_0y.$$

Thus if $c_0 \neq 0$, y can be determined so that $|a_0| \leqq \frac{1}{2}n|c_0|$. We can conclude therefore that either

$$(1.2) \qquad c_0 = 0$$

or integers x, y can be determined so that $|a_0| \leqq \frac{1}{4}n|a| \leqq |a|$. If $c_0 = 0$, then $W^{nx}A$ is a power of S and so $A \in G_{1,n}$. If $c_0 \neq 0$, then $|a_0| \leqq |a|$. If $|a_0| = |a|$, then the inequalities

$$|c_0| \leqq \tfrac{1}{2}|a| = \tfrac{1}{2}|a_0| \leqq \tfrac{1}{4}n|c_0| \leqq |c_0|$$

imply $a_0 = \pm 2$, $c_0 = \pm 1$, $n = 4$. But this is impossible, since $(a_0, nc_0) = 1$. Thus if $c_0 \neq 0$, then $|a_0| < |a|$. Therefore the process can be repeated until (1.2) holds. Thus $A \in G_{1,n}$, $\Gamma_0(n) \subset G_{1,n}$ and so $\Gamma_0(n) = G_{1,n}$.

Finally, put

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Then it is easy to verfy that

$$G_{2,2} = D^{-1}G_{1,4}D = D^{-1}\Gamma_0(4)D = \Gamma[2].$$

This completes the proof of the theorem.

Remark. We have implicitly identified a matrix with its negative in this theorem, but we keep the distinction elsewhere.

An interesting application of Theorem 1.1 is

THEOREM 1.2. *Let* α, β *be positive and suppose* G *is a group of type* (α,β) *with* $\alpha\beta \leqq 4$. *Then* G *is of finite index in* Γ.

It is not immediately obvious that there are non-abelian groups of type $(0,0)$. An example is furnished by the following theorem:

THEOREM 1.3. *There are non-abelian subgroups of* Γ *of type* $(0,0)$.

Proof. Put

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

and consider $[A, T]$. Since A is not of period 2 and $AT = TA^{-1}$, $[A, T]$ is non-abelian. Since $T^2 = -I$, every element of $[AT]$ is of one of the forms $\pm A^k$, $\pm TA^k$, k an integer. But A^k contains a zero entry if and only if $k = 0$, so that $\pm I$, $\pm T$ are the only elements of $[A, T]$ containing zero entries. This proves the theorem.

The previous construction can be generalized. It is not difficult to show that if A and B are elements of Γ such that $AB = BA^{-1}$ and A is not scalar, then both B and AB are similar to T over Γ. Further, essentially the same conclusion can be reached for $AB = BC$, if it is assumed that C is a polynomial in A.

We assume from now on that G is of type (α, β) with $\alpha\beta \neq 0$. We are interested in proving inclusion theorems for the groups $G(n)$ and for certain other groups, and we require the following result:

LEMMA 1.1. *Suppose that* m, n *are positive integers and that* $(m, \alpha\beta) = 1$. *Then* S^α, $W^{n\beta}$ *generate the coset representatives (left or right) of* $G(n)$ modulo $G(mn)$.

Proof. Let

$$M = \begin{bmatrix} a & b \\ nc & d \end{bmatrix} \in G(n).$$

It suffices to show that integers x, y can be found such that $W^{xn\beta}S^{y\alpha}M \in G(mn)$. Consider

$$S^{y\alpha}M = \begin{bmatrix} a + \alpha ncy & * \\ nc & * \end{bmatrix}.$$

We have $\Delta = (a, \alpha nc) = (a, \alpha)$ since M is unimodular. Thus $\Delta \mid \alpha$ and since $(\alpha, m) = 1$, $(\Delta, m) = 1$ also. Thus y may be found so that $(a + \alpha ncy, m) = 1$. For such a y, put $a_0 = a + \alpha ncy$. Then

$$W^{xn\beta}S^{y\alpha}M = \begin{bmatrix} * & * \\ n(\beta a_0 x + c) & * \end{bmatrix}.$$

Since $(\beta a_0, m) = 1$, x can be determined so that $\beta a_0 x + c \equiv 0 \pmod{m}$. For this choice of x, $W^{xn\beta} S^{y\alpha} M \in G(mn)$ and the proof of the lemma is complete.

Since x and y were determined only modulo m it is clear that $(G(n): G(mn)) < \infty$.

On the basis of this lemma and formula (1.1), the theorem that follows can be proved in just the same way that the corresponding theorem for Γ was proved in [8].

THEOREM 1.4. *Let m and n be positive integers, and assume that $(m, \alpha\beta) = 1$. If H is any group such that*

$$G(mn) \subset H \subset G(n),$$

then $H = G(dn)$ for some divisor d of m.

This theorem can be generalized to a larger class of groups $G(m, n)$ defined as follows: The element

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

of G belongs to $G(m, n)$ if and only if $b \equiv 0 \pmod{m}$, $c \equiv 0 \pmod{n}$. Thus $G(n) = G(1, n)$. We have as a matter of fact

THEOREM 1.5. *Let m and n be positive integers and assume that $(m, n\alpha\beta) = 1$. Let H be any subgroup of G containing $G(m, n)$. Then $H = G(m_1, n_1)$ for some divisors m_1 of m and n_1 of n.*

Proof. Since $(m, n\alpha\beta) = 1$, we can determine integers x, y such that $xm - yn\alpha\beta = 1$. Set

$$R = \begin{bmatrix} xm & \alpha \\ yn\beta & 1 \end{bmatrix}.$$

Then $R \in G$ since $R = S^{\alpha} W^{yn\beta}$. The proof is now completed in the same way as the proof of the corersponding theorem given in [9] for Γ.

For purposes of constructing modular functions we need more detailed information on the structure of the representatives of $G(n)$ modulo $G(mn)$ than is furnished by Lemma 1.1. This is given in the lemmas below:

LEMMA 1.2. *Let m and n be positive integers such that m is divisible only by primes dividing n and $(n, \beta) = 1$. Then $(G(n): G(mn)) = m$, and*

$$W^{xn\beta}, \qquad 0 \leqq x \leqq m - 1, \tag{1.3}$$

is a set of left coset representatives for $G(n)$ modulo *$G(mn)$.*

Proof. Let

$$M = \begin{bmatrix} a & b \\ nc & d \end{bmatrix} \in G(n).$$

Consider

$$W^{-xn\beta}M = \begin{bmatrix} * & * \\ n(c - xa\beta) & * \end{bmatrix}.$$

Now $(a\beta, n) = 1$ since $\det M = 1$ and $(\beta, n) = 1$ by assumption. Thus $(a\beta, m) = 1$ also, since m is divisible only by primes dividing n. Hence we can find x such that $xa\beta \equiv c \pmod{m}$, $0 \leqq x \leqq m - 1$. For this x, $M = W^{xn\beta}M_0$, where $M_0 \in G(mn)$. Further, if $0 \leqq x, y \leqq m - 1$, then $W^{-xn\beta}W^{yn\beta} \in G(mn)$ if and only if $n\beta(y - x) \equiv 0 \pmod{mn}$, which implies that $x = y$ since $(\beta, m) = 1$. The representatives (1.3) are thus distinct modulo $G(mn)$ and the proof of the lemma is completed.

LEMMA 1.3. *Let m and n be positive integers such that m is a power of a prime p and $(p, na\beta) = 1$. Set $m' = m/p$. Choose t so that $na\beta t \equiv -1 \pmod{p}$ and put $R = S^{\alpha}W^{nt\beta}$. Then $(G(n) : G(mn)) = m + m'$ and*

$$W^{xn\beta}, \quad 0 \leqq x \leqq m - 1, \quad RW^{ynp\beta}, \quad 0 \leqq y \leqq m' - 1, \tag{1.4}$$

is a set of left coset representatives for $G(n)$ modulo $G(mn)$.

Proof. Let

$$M = \begin{bmatrix} a & b \\ nc & d \end{bmatrix} \in G(n).$$

Suppose first that $(a, p) = 1$. Consider

$$W^{-xn\beta}M = \begin{bmatrix} * & * \\ n(c - xa\beta) & * \end{bmatrix}.$$

Since $(a\beta, p) = 1$ and m is a power of p, we can find x so that $xa\beta \equiv c \pmod{m}$, $0 \leqq x \leqq m - 1$. For this x, $M = W^{xn\beta}M_0$, where $M_0 \in G(mn)$. Suppose now that $p \mid a$. Consider

$$W^{-xn\beta}S^{-\alpha}M = \begin{bmatrix} * & * \\ n(c - x\beta(a - \alpha nc)) & * \end{bmatrix}.$$

Now $(\beta(a - \alpha nc), p) = (\alpha\beta nc, p) = 1$ since $p \mid a$, $\det M = 1$, and $(n\alpha\beta, p) = 1$ by assumption. Thus x can be determined modulo m so that

$$x\beta(a - \alpha nc) \equiv c \pmod{m}. \tag{1.5}$$

For this x, $M = S^{\alpha}W^{xn\beta}M_0$, where $M_0 \in G(mn)$. Considering congruence (1.5) modulo p, we have $n\alpha\beta x \equiv -1 \pmod{p}$. Let t satisfy $n\alpha\beta t \equiv -1$

$(\mathrm{mod}\, p)$. Then $x = t + yp$, where now y is determined modulo m', and $S^\alpha W^{xn\beta}$ becomes $RW^{ynp\beta}$.

Further, suppose that $0 \leqq x, x' \leqq m - 1$ and $0 \leqq y, y' \leqq m' - 1$. Then $W^{-xn\beta}W^{x'n\beta} \in G(mn)$ if and only if $n\beta(x' - x) \equiv 0 \ (\mathrm{mod}\, mn)$, or $x = x'$; $(RW^{ynp\beta})^{-1}RW^{y'np\beta} \in G(mn)$ if and only if $np\beta(y' - y) \equiv 0 \ (\mathrm{mod}\, mn)$, or $y = y'$; and finally, $W^{-xn\beta}RW^{ynp\beta} \in G(mn)$ implies that $t \equiv 0 \ (\mathrm{mod}\, p)$, which is impossible. The representatives (1.4) are therefore distinct modulo $G(mn)$, completing the proof of the lemma.

The application to the construction of modular functions we wish to make depends on these lemmas and Theorem 2.2 of [7]. We obtain

THEOREM 1.6. *Let m be a power of a prime p, and set $m' = m/p$. Suppose that $h(\tau)$ is a function on $G(m)$. Define*

$$f_m(\tau) = \sum_{k=0}^{m-1} h(W^{-k\beta}\tau),$$

$$g_m(\tau) = \sum_{k=0}^{m-1} h(W^{-kp\beta}\tau).$$

Then if $(p, \beta) = 1$, $g_m(\tau)$ is a function on $G(p)$ and if $(p, \alpha\beta) = 1$, $f_m(\tau) + g_{m'}(R^{-1}\tau)$ is a function on G, where $R = S^\alpha W^{t\beta}$ and $\alpha\beta t \equiv -1$ $(\mathrm{mod}\, p)$.

Since R is in G, we can equally well conclude that in the latter case $f_m(R\tau) + g_{m'}(\tau)$ is a function on G, and since $g_{m'}(\tau)$ is a function on $G(p)$, $f_m(R\tau)$ is a function on $G(p)$ also.

Define the subgroup

$$K = [T, S^2].$$

Then K is of type $(2, 2)$ and of index 3 in Γ. In general $[T, S^n]$ is of type (n, n) and of infinite index for $n \geqq 3$. For $n = 1$ it is just Γ. K can also be described as follows: The element A of Γ is in K if and only if $A \equiv I$ $(\mathrm{mod}\, 2)$ or $A \equiv T \ (\mathrm{mod}\, 2)$. From Lemmas 1.2 and 1.3 we obtain for K

LEMMA 1.4. *Let p be an odd prime, m a power of p, and set $m' = m/p$. Then $(K(p) : K(m)) = m'$ and*

$$W^{2pk}, \qquad 0 \leqq k \leqq m' - 1, \tag{1.6}$$

is a set of left coset representatives for $K(p)$ modulo $K(m)$.

LEMMA 1.5. *Let p be an odd prime, m a power of p, and set $m' = m/p$. Put $R = S^2 W^{\frac{1}{2}(p^2-1)}$. Then $(K : K(m)) = m + m'$ and*

(1.7) $\qquad W^{2k},\ 0 \leqq k \leqq m-1,\ RW^{2pk},\ 0 \leqq k \leqq m'-1,$

is a set of left coset representatives for K modulo $K(m)$.

The group K is the underlying group for the study of the modular form whose powers generate $r_s(n)$. This study will be undertaken in § 3.

Let G be an arbitrary subgroup of Γ and let C be a fixed element of G, n a positive integer. Define the congruence subgroup $G(n;C)$ as follows: The element A of G belongs to $G(n;C)$ if and only if for same integer k, $A \equiv C^k \pmod{n}$.

These groups are congruence subgroups and have an easy structure. Thus $G(n;I)$ is the principal congruence subgroup of G of level n, and if e is the exponent of C modulo n, then $(G(n;C):G(n;I)) = e$, the quotient group $G(n;C)/G(n;I)$ being isomorphic to the cyclic group generated by C modulo n. We remark that a principal congruence subgroup of G is normal in any subgroup of G containing it, and a subgroup of G containing a principal congruence subgroup is itself a congruence subgroup of G.

Let L be the complex of elements $A \in G(n;C)$ such that $A \equiv C \pmod{n}$. The following lemma is of use in showing a function invariant for $G(n;C)$:

LEMMA 1.6. *If $g(\tau)$ is invariant for L, then $g(\tau)$ is a function on $G(n;C)$.*

Proof. Since $C \in L$, $g(C\tau) = g(\tau)$; and by iteration, $g(\tau)$ is invariant for any power of C. Let $A \in G(n;C)$. Then for some integer k, $A \equiv C^k \pmod{n}$. Hence $C^{1-k}A \in L$ and so $g(\tau)$ is invariant for $C^{1-k}A$. Since $g(\tau)$ is invariant for C^{1-k}, $g(\tau)$ is invariant for A. The lemma is thus proved.

We note the facts

$$K = \Gamma(2;T), \qquad K(n) = \Gamma^*(n)(2;T),\ n \text{ odd}.$$

2. The modular form $\vartheta(\tau)$. The functions to be considered in the following section will be built up from the modular form $\vartheta(\tau)$, defined by

$$\vartheta(\tau) = \sum_{n=-\infty}^{\infty} q^{n^2}, \quad q = \exp \pi i\tau, \operatorname{im}\tau > 0. \tag{2.1}$$

Thus

$$\vartheta^s(\tau) = \sum_{n=0}^{\infty} r_s(n) q^n.$$

Here q is the proper uniformizing variable for $\vartheta(\tau)$ at the parabolic point $i\infty$ of the fundamental region of K, which will be denoted by K^*. A good discussion of uniformizing variables can be found in [2].

$\vartheta(\tau)$ can also be defined in terms of the Dedekind modular form $\eta(\tau)$, where

$$(2.2) \qquad \eta(\tau) = x^{1/24}\prod_{n=1}^{\infty}(1-x^n), \quad x = q^2 = \exp 2\pi i\tau, \operatorname{im}\tau > 0,$$

and x is the proper uniformizing variable for $\eta(\tau)$ at the parabolic point $i\infty$ of Γ^*. The relationship is

$$(2.3) \quad \vartheta(\tau) = \eta^2(\tfrac{1}{2}(\tau+1))\eta^{-1}(\tau+1) = \prod_{n=1}^{\infty}(1-(-q)^n)(1+(-q)^n)^{-1}.$$

Both $\vartheta(\tau)$ and $\eta(\tau)$ are of dimension $-\frac{1}{2}$ and are regular and zero-free in the interior of the upper τ half-plane. Thus $\vartheta(\tau)$ is an entire modular form for K and $\eta(\tau)$ is an entire modular form for Γ.

The transformation equations for these functions are classical and have been worked out in great detail by many writers. In particular,

$$(2.4) \qquad \vartheta(S^2\tau) = \vartheta(\tau),$$

$$(2.5) \qquad \vartheta(T\tau) = (-i\tau)^{\frac{1}{2}}\vartheta(\tau)$$

and

$$(2.6) \qquad \vartheta(W\tau) = 2(\tau+1)^{\frac{1}{2}}x^{1/8}\prod_{n=1}^{\infty}(1-x^{2n})(1+x^n).$$

As (2.6) indicates, x is the proper uniformizing variable for $\vartheta(\tau)$ at the parabolic point 1 of K^*, since W takes $i\infty$ into 1.

In his paper [4] Hecke proves the transformation formula below, which we state as a lemma:

Lemma 2.1. (*Hecke*, [4]). *Let*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in K, A \equiv T \pmod 2.$$

Then

$$(2.7) \qquad \vartheta(A\tau) = (-i(c\tau+d))^{\frac{1}{2}} i^{\frac{1}{2}(1-c)}(a/|c|)\vartheta(\tau),$$

where $(a/|c|)$ *is the generalized Legendre-Jacobi symbol of quadratic reciprocity.*

To obtain the transformation formula for a matrix $A \in K$ such that $A \equiv I \pmod 2$, we write $A = T^{-1}TA$ and apply (2.5) and (2.7). In this manner we obtain

Lemma 2.2. *Let*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in K, A \equiv I \pmod 2.$$

Then

$$\vartheta(A\tau) = (-i(c\tau+d))^{\frac{1}{2}} i^{\frac{1}{2}(1-a)} (-c/|a|)\vartheta(\tau). \tag{2.8}$$

3. Sums of squares. We assume in this section that p is an odd prime and that m is an even power of p. We set $m' = m/p$. We prove first

THEOREM 3.1. *Let s be an integer, and define*

$$h(\tau) = h(s, p, m; \tau) = \vartheta^s(m\tau)\vartheta^{-s}(\tau). \tag{3.1}$$

Then $h(\tau)$ is an entire modular function on $K(m)$.

Proof. Suppose

$$A = \begin{bmatrix} a & b \\ mc & d \end{bmatrix} \in K(m).$$

Put

$$A_0 = \begin{bmatrix} a & mb \\ c & d \end{bmatrix}.$$

Then $A_0 \in K$, $A \equiv A_0 \pmod 2$, and we have

$$h(A\tau) = \vartheta^s(A_0 m\tau)\vartheta^{-s}(A\tau).$$

Since $K(m) = \Gamma_0(m)(2; T)$, we need only consider $A \equiv T \pmod 2$, in view of Lemma 1.6. Then Lemma 2.1 implies that

$$h(A\tau) = i^{\frac{1}{2}(m-1)cs}(a/m)^s h(\tau),$$

and this becomes $h(\tau)$ since m is an odd square. Thus $h(\tau)$ is a function on $K(m)$. The fact that $h(\tau)$ is an *entire* function on $K(m)$ is clear, since $\vartheta(\tau)$ is regular and zero-free in the interior of the upper τ half-plane.

We now want to construct a class of entire modular functions on K and we do this by Theorem 1.6, Lemmas 1.4 and 1.5 and the preceding theorem. we obtain

THEOREM 3.2. *Put*

$$f_m(\tau) = f(s, p, m; \tau) = \sum_{k=0}^{m-1} h(W^{-2k}R\tau), \tag{3.2}$$

$$g_m(\tau) = g(s, p, m; \tau) = \sum_{k=0}^{m-1} h(W^{-2pk}\tau), \tag{3.3}$$

where $h(\tau)$ is defined by (3.1) and $R = S^2 W^{\frac{1}{2}(p^2-1)}$. Then $f_m(\tau)$, $g_{m'}(\tau)$, $g_m(\tau)$ are entire modular functions on $K(p)$ and

$$F_m(\tau) = f_m(\tau) + g_{m'}(\tau) \tag{3.4}$$

is an entire modular function on K.

Our general procedure will be as follows: K^*, which is of genus zero, has the two inequivalent parabolic points $\tau=i\infty$, $\tau=1$. The functions $F_m(\tau)$ are entire modular functions on K with polar singularities at most in the proper uniformizing variables (q for $i\infty$, x for 1) at these parabolic points. Therefore we need only determine the behaviour of these functions at $i\infty$ and 1 to have complete information about them. We shall show that $F_m(\tau)$ is pole-free at $\tau=i\infty$, and for positive s has a pole at $\tau=1$ of order less than $s/8$, a number *independent of* m. This will imply that for sufficiently many different values of m the functions $F_m(\tau)$ are linearly dependent, if $s>0$. Comparing coefficients we will obtain the desired identities for $r_s(n)$.

We consider $\tau=i\infty$ first. The function $f_m(\tau)$ can be simplified by noticing that

$$R^{-1}T=\begin{bmatrix}-2 & -1\\ p^2 & \tfrac{1}{2}(p^2-1)\end{bmatrix}\in K(p);$$

and since $f_m(\tau)$ is a function on $K(p)$, we have

$$\begin{aligned} f_m(\tau) &= f_m(R^{-1}T\tau)\\ &= \sum_{k=0}^{m-1} h(W^{-2k}T\tau)\\ &= \sum_{k=0}^{m-1} h(TS^{2k}\tau)\\ &= \sum_{k=0}^{m-1} \vartheta^s(T(\tau+2k)/m)\,\vartheta^{-s}(T(\tau+2k)). \end{aligned}$$

The transformation formula (2.5) may now be employed to show that

$$f_m(\tau)=m^{-\frac{1}{2}s}\vartheta^{-s}(\tau)\sum_{k=0}^{m-1}\vartheta^s((\tau+2k)/m). \tag{3.5}$$

Therefore $f_m(\tau)$ is pole-free at $\tau=i\infty$ since $\vartheta(\tau)$ is both pole and zero-free at $\tau=i\infty$.

We now consider $g_{m'}(\tau)$ at $\tau=i\infty$. We have

$$\begin{aligned} g_{m'}(\tau) &= \sum_{k=0}^{m'-1} h(W^{-2pk}\tau)\\ &= \sum_{k=0}^{m'-1} \vartheta^s(mW^{-2pk}\tau)\,\vartheta^{-s}(W^{-2pk}\tau)\\ &= \sum_{k=0}^{m'-1} \vartheta^s(m'W^{-2k}p\tau)\,\vartheta^{-s}(W^{-2pk}\tau). \end{aligned}$$

Let $(m',k)=\Delta$ and put $m'=\Delta m_0$, $k=\Delta k_0$, where now $(m_0,k_0)=1$. Determine integers m_1, k_1 so that $m_0m_1-4k_0k_1=1$ and set

$$M = \begin{bmatrix} m_0 & -2k_1 \\ -2k_0 & m_1 \end{bmatrix}.$$

Then $M \in K$ and $M \equiv I \pmod 2$. Furthermore,

$$m' W^{-2k} p\tau = M(\Delta p\tau + 2k_1)/(m'/\Delta).$$

Now $W^{-2pk} \in K$, $W^{-2pk} \equiv I \pmod 2$ also. Thus using (2.8) we find that

$$h(W^{-2pk}\tau) = c\vartheta^s((\Delta p\tau + 2k)/(m'/\Delta))\vartheta^{-s}(\tau),$$

where c is a number independent of τ. Therefore $g_{m'}(\tau)$ is also pole-free at $\tau = i\infty$ and we have proved

LEMMA 3.1. *For all integral s, $F_m(\tau)$ is pole-free at $\tau = i\infty$.*

We go on now to the remaining parabolic point $\tau = 1$. As indicated, we study the function $F_m(W\tau)$ at $\tau = i\infty$ in terms of the uniformizing variable $x = \exp 2\pi i\tau$. As with $\tau = i\infty$, we consider each component $f_m(W\tau)$, $g_{m'}(W\tau)$ separately. Now

$$\begin{aligned} g_{m'}(W\tau) &= \sum_{k=0}^{m'-1} h(W^{-2pk+1}\tau) \\ &= \sum_{k=0}^{m'-1} \vartheta^s(mW^{-2pk+1}\tau)\vartheta^{-s}(W^{-2pk+1}\tau). \end{aligned}$$

But $(m, -2pk+1) = 1$ since m is power of p. Thus integers m_1, k_1 can be determined so that $mm_1 - 2(2pk-1)k_1 = 1$. Put

$$M = \begin{bmatrix} m & -2k_1 \\ 1-2pk & m_1 \end{bmatrix}.$$

Then

$$mW^{-2pk+1}\tau = M(\tau + 2k_1)/m.$$

Now $MW^{-1} \in K$, and $MW^{-1} \equiv I \pmod 2$. Using (2.8) once again, we find that for some number c independent of τ,

$$h(W^{-2pk+1}\tau) = c\vartheta^s(W(\tau + 2k_1)/m)\vartheta^{-s}(W\tau).$$

We can now use (2.6) to deduce that apart from constant multipliers, the latter expression begins with

$$x^{-\epsilon}, \epsilon = (s/8)(1 - 1/m).$$

From this we conclude that $g_{m'}(W\tau)$ is pole-free if $s \leqq 0$ and has a pole of order less than $s/8$ if $s > 0$.

Turning now to $f_m(W\tau)$ and making use of (3.5) we have that

$$f_m(W_T) = m^{-\frac{1}{2}s}\vartheta^{-s}(W_T)\sum_{k=0}^{m-1}\vartheta^s((1/m)S^{2k}W_T).$$

Put $\Delta = (2k+1, m)$. Set $2k+1 = \Delta k_0, m = \Delta m_0$ so that $(m_0, k_0) = 1$. Determine k_1, m_1 so that $2k_0k_1 - m_0m_1 = 1$. Put

$$M = \begin{bmatrix} k_0 & m_1 \\ m_0 & 2k_1 \end{bmatrix}.$$

Then

$$(1/m)S^{2k}W_T = M(\Delta\tau + \Delta - 2k_1)/(m/\Delta).$$

Furthermore, $MW^{-1} \in K$ and $MW^{-1} \equiv T \pmod 2$. Using (2.7) and then (2.6), we find that

$$\vartheta^s((1/m)S^{2k}W_T)\vartheta^{-s}(W_T)$$

begins with

$$x^{-\epsilon}, \epsilon = (s/8)(1 - \Delta^2/m),$$

apart from constant multipliers. Considering $s \leqq 0$ and $s > 0$ separately, we conclude easily

LEMMA 3.2. *If $s \leqq 0$, then $F_m(\tau)$ has a pole of order $-s(m^2-1)/8$ at most at $\tau = 1$. If $s > 0$, then $F_m(\tau)$ has a pole of order less than $s/8$ at $\tau = 1$.*

Summarizing for $s > 0$, Lemmas 3.1 and 3.2 taken together state

THEOREM 3.3. *Suppose $s > 0$. Then the total valence of $F_m(\tau)$ throughout K^* is less than $s/8$.*

From this theorem we deduce the following one by a simple application of Liouville's theorem:

THEOREM 3.4. (*First principal theorem*). *Let μ be any integer satisfying*

$$\mu \geqq s/8, \qquad s > 0.$$

Then if m_k, $1 \leqq k \leqq \mu$, are all even powers of p there exist constants C_k, $0 \leqq k \leqq \mu$, not all zero such that

$$\sum_{k=1}^{\mu} C_k F_{m_k}(\tau) = C_0. \tag{3.6}$$

Identity (3.6) is the identity whose existence we wished to prove. We are now interested is comparing coefficients in this identity to obtain identies for $r_s(n)$, but first we must know the Fourier expansion of $F_m(\tau)$ at $\tau = i\infty$.

The component $f_m(\tau)$ offers no difficulty whatsoever and we find

$$f_m(\tau) = m^{1-\frac{1}{2}s}\vartheta^{-s}(\tau)\sum_{n=0}^{\infty} r_s(mn)\,q^n. \tag{3.7}$$

$g_{m'}(\tau)$ is quite complicated however and we must resort to special devices.

We write $k:n$ in a sumation to indicate that k runs over a reduced set of residues modulo n. Then if ϕ is an arbitrary function and n is a power of p, it is evident that

$$\sum_{k=0}^{n-1}\phi(k) = \sum_{k:n}\phi(k) + \sum_{k=0}^{n'-1}\phi(pk), \tag{3.8}$$

where $n' = n/p$. Iterating, we can prove without difficulty

LEMMA 3.3. *If $n = p^t$, then*

$$\sum_{k=0}^{n-1}\phi(k) = \phi(0) + \sum_{r=1}^{t}\sum_{k:p^r}\phi(kp^{t-r}). \tag{3.9}$$

If we apply Lemma 3.3 to $g_{m'}(\tau)$, we find

LEMMA 3.4. *Suppose $m = p^{2t}$. Then*

$$g_{m'}(\tau) = h(\tau) + \sum_{r=1}^{2t-1}\phi_r(\tau), \tag{3.10}$$

where

$$\phi_r(\tau) = \sum_{k:p^r} h(W^{-2p^{2t-r}k}\tau). \tag{3.11}$$

In this form we can expand $g_{m'}(\tau)$. We have

$$\begin{aligned} h(W^{-2p^{2t-r}k}\tau) &= \vartheta^s(p^{2t}W^{-2p^{2t-r}k}\tau)\,\vartheta^{-s}(W^{-2p^{2t-r}k}\tau) \\ &= \vartheta^s(p^r W^{-2k}p^{2t-r}\tau)\,\vartheta^{-s}(W^{-2p^{2t-r}k}\tau). \end{aligned}$$

Now $(p^r, k) = 1$ since k runs over a reduced set of residues modulo p^r. Thus integers m_1, k_1 can be determined so that $p^r m_1 - 4kk_1 = 1$. Put

$$M = \begin{bmatrix} p^r & -2k_1 \\ -2k & m_1 \end{bmatrix}.$$

Then $M \in K$, $M \equiv I \pmod 2$. Also

$$p^r W^{-2k}p^{2t-r}\tau = M(p^{2t-r}\tau + 2k_1)/p^r.$$

We can now employ (2.8) to simplify $h(W^{-2p^{2t-r}k}\tau)$. The result is that

$$h(W^{-2p^{2t-r}k}\tau) = c\vartheta^s((p^{2t-r}\tau + 2k_1)/p^r)\,\vartheta^{-s}(\tau),$$

where

$$c = p^{-\frac{1}{2}rs} i^{\frac{1}{4}s(1-p^r)} (-2k_1/p)^{rs}.$$

Noting now that k and k_1 simultaneously run over a reduced set of residues modulo p^r, we find

$$g_{m'}(\tau) = h(\tau) + \vartheta^{-s}(\tau) \sum_{r=1}^{2t-1} p^{-\frac{1}{2}rs} i^{\frac{1}{4}s(1-p^r)} (-2/p)^{rs} G_r(\tau),$$

where

$$G_r(\tau) = \sum_{k:p^r} (k/p)^{rs} \vartheta^s((p^{2t-r}\tau + 2k)/p^r).$$

The expansion of $G_r(\tau)$ is straightforward and depends only on the Gauss sums. We find that

$$(3.12) \qquad G_r(\tau) = p^{r-1} \sum_{n=0}^{\infty} \sum_{k:p} (k/p)^{rs} e^{2\pi ink/p} r_s(p^{r-1}n) q^{p^{2t-1-r}n}.$$

Let us define

$$(3.13) \qquad \gamma_{rs}(n) = \begin{cases} \sum_{k:p} (k/p)^{rs} e^{2\pi ink/p} & n \text{ an integer,} \\ 0 & \text{otherwise} \end{cases}$$

so that for integral n

$$\gamma_{rs}(n) = \begin{cases} p-1 & rs \text{ even, } p \mid n \\ -1 & rs \text{ even, } (n,p) = 1 \\ i^{(\frac{1}{2}(p-1))^2} (n/p) p^{\frac{1}{2}} & rs \text{ odd.} \end{cases}$$

Then (3.12) may be written

$$G_r(\tau) = p^{r-1} \sum_{n=0}^{\infty} \gamma_{rs}(p^{r+1-2t}n) r_s(p^{2r-2t}n) q^n.$$

We thus obtain

LEMMA 3.5. *Define*

$$C_{r,s} = p^{r-\frac{1}{2}rs-1} i^{\frac{1}{4}s(1-p^r)} (-2/p)^{rs}.$$

Then

$$(3.14) \quad g_{m'}(\tau) = \vartheta^{-s}(\tau) \sum_{n=0}^{\infty} \{ r_s(n/m) + \sum_{r=1}^{2t-1} C_{r,s} \gamma_{rs}(p^{r+1-2t}n) r_s(p^{2r-2t}n) \} q^n.$$

Expansion (3.14) now yields our second principal theorem.

THEOREM 3.5. (*Second principal theorem*). *Let* μ *be any integer satisfying*

$$\mu \geqq s/8, \quad s > 0.$$

choose $m_k = p^{2t_k}$, $1 \leqq k \leqq \mu$. *Then there are numbers* c_k, $0 \leqq k \leqq \mu$, *not all zero and independent of* n *such that for all integral* n

$$\begin{aligned}(3.15)\quad & \sum_{k=1}^{\mu} C_k \{ m_k^{1-\frac{1}{2}s} r_s(m_k n) + r_s(n/m_k) \\ & \qquad + \sum_{r=1}^{2t_k-1} C_{r,s}\gamma_{rs}(p^{r+1-2t_k} n) r_s(p^{2r-2t_k} n) \} = C_0 r_s(n).\end{aligned}$$

The simplest examples (and the only ones we shall write down explicitly) occur for $s \leqq 8$. For then the choice $\mu = 1$ is permissible, and choosing $m = p^2$, we find

$$\begin{aligned}(3.16)\quad r_s(np^2) &= \{1 + p^{s-2} - (-1)^{\frac{1}{4}(s-1)(p-1)} p^{\frac{1}{2}(s-3)} (n/p)\} r_s(n) \\ &\quad - p^{s-2} r_s(n/p^2), \qquad s = 1, 3, 5, 7,\end{aligned}$$

$$\begin{aligned}(3.17)\quad r_s(np^2) &= \{1 + p^{s-2} + (-1)^{\frac{1}{4}s(p-1)} p^{\frac{1}{2}(s-2)} (n/p)^2\} r_s(n) \\ &\quad - p^{s-2} r_s(n/p^2), \qquad s = 2, 4, 6, 8.\end{aligned}$$

The identity for $s = 3$ can be found in [11]. See also the paper [5] where the identities for $s = 3, 5, 7$ are derived by the application of a Hecke operator.

As a final application, we notice that the entire modular function $F(-1, 3, 9; \tau)$ is non-constant, and is regular throughout K^* except at $\tau = 1$ where it has a pole of order 1. Therefore

THEOREM 3.6. *The function* $F(-1, 3, 9; \tau)$ *is a Hauptmodul for* K.

NATIONAL BUREAU OF STANDARDS,
WASHINGTON, D. C.

REFERENCES.

[1] J. L. Brenner, "Quelques groupes libres de matrices," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 241 (1955), pp. 1689-1691.

[2] L. R. Ford, *Automorphic Functions*, New York, 1951.

[3] K. Goldberg and M. Newman, "Pairs of matrices of order two which generate free groups," *Illinois Journal of Mathematics*, vol. 1 (1957), pp. 446-448.

[4] E. Hecke, "Herleitung des Euler-Produktes der Zetafunktion und einiger L-reihen aus ihrer Funktionalgleichung," *Mathematische Annalen*, vol. 119 (1944), pp. 266-287.

[5] J. H. Lint, *Hecke operators and Euler products*, Dissertation, University of Utrecht, 1957.

[6] M. Newman, "On the existence of identities for the coefficients of certain modular forms," *The Journal of the London Mathematical Society*, vol. 30 (1955), pp. 488-493.

[7] ———, "Remarks on some modular identities," *Transactions of the Mathematical Society*, vol. 73 (1952), pp. 313-320.

[8] ———, "Structure theorems for modular subgroups," *Duke Mathematical Journal*, vol. 22 (1955), pp. 25-32.

[9] ———, "An inclusion theorem for modular groups," *Proceedings of the American Mathematical Society*, vol. 8 (1957), pp. 125-127.

[10] ———, "Modular forms whose coefficients possess multiplicative properties," *Annals of Mathematics*, vol. 70 (1959), pp. 478-489.

[11] G. Pall, "On the arithmetic of quaternions," *Transactions of the American Mathematical Society*, vol. 47 (1940), pp. 487-500.

ABELIAN EXTENSIONS OF DIFFERENTIAL FIELDS.*

By E. R. Kolchin.[1]

Introduction. The Galois theory of differential fields ([2], [3]) is concerned with strongly normal extensions of a differential field $\mathcal{F}$ with algebraically closed field of constants $\mathcal{C}$ of characteristic 0. The Galois group $\mathfrak{G}$ of such an extension $\mathcal{G}$ has a sort of algebraico-geometric structure and is shown to be birationally isomorphic with an algebraic group variety (not necessarily connected); more precisely, there is an algebraic group G in the algebraic geometry having for universal domain the fixed universal differential field extension $\mathcal{U}$ of $\mathcal{G}$, G being defined over $\mathcal{C}$, such that $\mathfrak{G}$ is birationally isomorphic with the group $G_{\mathcal{K}}$ consisting of the points of G which are rational over the field of constants $\mathcal{K}$ of $\mathcal{U}$ ($G_{\mathcal{K}}$ is an algebraic group in the algebraic geometry with universal domain $\mathcal{K}$). This permits the classification of strongly normal extensions by means of algebraic groups: a strongly normal extension of $\mathcal{F}$ with Galois group birationally isomorphic to the subgroup $G_{\mathcal{K}}$ of an algebraic group G defined over $\mathcal{C}$, or to an algebraic subgroup of $G_{\mathcal{K}}$, will be called a *G-extension* of $\mathcal{F}$.

If $\mathcal{G}$ is a G-extension of $\mathcal{F}$ and if

$$G = G_0 \supset G_1 \supset \cdots \supset G_r = 1$$

is a normal chain of algebraic subgroups of G, all defined over $\mathcal{C}$, then there is a corresponding tower

$$\mathcal{F} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_r = \mathcal{G}$$

of differential fields such that $\mathcal{F}_i$ is a G_{i-1}/G_i-extension of $\mathcal{F}_{i-1}$ $(1 \leqq i \leqq r)$. Now, by Chevalley's structure theorem (see Rosenlicht [4], p. 439, or Barsotti [1], p. 47) G has a canonical normal chain

$$G \supset G^0 \supset H^0 \supset 1 \tag{1}$$

defined over $\mathcal{C}$, G^0 being the component of the unity element of G and H^0 being the greatest connected linear algebraic subgroup of G^0, such that the

* Received October 26, 1959.

[1] This work was done in connection with a contract with the Office of Naval Research.

factor groups G/G^0 and G^0/H^0 are, respectively, finite and abelian (i. e. an abelian variety). Therefore $\mathcal{F}$ and $\mathcal{G}$ appear in a corresponding tower

$$\mathcal{F} \subset \mathcal{F}^0 \subset \mathcal{E}^0 \subset \mathcal{G}. \tag{2}$$

But a finite group corresponds to an extension of finite degree, so that $\mathcal{F}^0$ is finite algebraic over $\mathcal{F}$ (indeed, $\mathcal{F}^0$ is the algebraic closure of $\mathcal{F}$ in $\mathcal{G}$), and ([3] p. 891) a linear group corresponds to a Picard-Vessiot extension, so that $\mathcal{G}$ is Picard-Vessiot over $\mathcal{E}^0$. Thus, the gap in our knowledge of the structure of $\mathcal{G}$ over $\mathcal{F}$ is in the extension from $\mathcal{F}^0$ to $\mathcal{E}^0$. The purpose of the present paper is to close this gap, that is, to characterize the strongly normal extensions which are A-extensions for some abelian variety A. We shall call any such extension an *abelian* extension. The characterization which we find is in terms of solutions of certain kinds of differential equations, and generalizes the result previously obtained ([2] p. 809, [3] p. 892) in the case in which the extension is of transcendence degree 1, i. e. A is of dimension 1. The remainder of the present introduction is intended to describe this characterization more fully, in its natural more general setting which includes, among others, the above mentioned characterization of strongly normal extensions with linear groups.

Let G be any *connected* algebraic group defined over $\mathcal{K}$, of dimension say n. If ω is a differential on G and α is a point of G at which ω is finite, there is induced a differential $\omega(\alpha)$ of $\mathcal{U}$ over $\mathcal{K}$ (if $\omega = df$ and the rational function f on G is defined over $\mathcal{K}$ then $\omega(\alpha)$ maps each derivation $\mathfrak{d}$ of $\mathcal{U}$ over $\mathcal{K}$ onto $\mathfrak{d}f(\alpha)$). Every birational automorphism ϕ of G induces a contravariant automorphism ϕ^* of the vector space $\mathfrak{I}(G)$ over $\mathcal{U}$ of *invariant* (i. e. left invariant) differentials on G; in particular, the inner automorphism $\tau(\beta)$: $s \rightarrow \beta s \beta^{-1}$ determined by an element β of G determines an automorphism $\tau(\beta)^*$ of $\mathfrak{I}(G)$. It is "well-known" that if ω is invariant, and therefore finite at every point of G, then

$$\omega(\alpha\beta) = (\tau(\beta^{-1})^*\omega)(\alpha) + \omega(\beta) \tag{3}$$

for all points α and β of G. This formula (which when G is the additive or multiplicative group of $\mathcal{U}$ becomes, respectively, $\mathfrak{d}(\alpha+\beta) = \mathfrak{d}\alpha + \mathfrak{d}\beta$ or $(\alpha\beta)^{-1}\mathfrak{d}(\alpha\beta) = \alpha^{-1}\mathfrak{d}\alpha + \beta^{-1}\mathfrak{d}\beta$, $\mathfrak{d}$ representing an arbitrary derivation of $\mathcal{U}$ over $\mathcal{K}$, and when G is the general linear group $GL(r)$ becomes $(\alpha\beta)^{-1}\mathfrak{d}(\alpha\beta) = \beta^{-1}(\alpha^{-1}\mathfrak{d}\alpha)\beta + \beta^{-1}\mathfrak{d}\beta$) has been given a rigorous exposition by Rosenlicht [5], but unfortunately only in the case of commutative groups G; therefore we give in § 1 a discussion supplementing that of Rosenlicht, establishing the desired formula in full generality.

We define a point α of G to be a *G-primitive* over $\mathfrak{F}$ if the following two conditions are satisfied: 1° the field of constants of $\mathfrak{F}\langle\alpha\rangle$ is $\mathfrak{K}$; 2° $\langle\delta,\omega(\alpha)\rangle\in\mathfrak{F}$ for every derivation operator δ of the differential field and every invariant differential ω on G defined over $\mathfrak{K}$. If the derivation operators are denoted by $\delta_1,\cdots,\delta_m$ and if n linearly independent differentials $\omega_1,\cdots,\omega_n$ defined over $\mathfrak{K}$ are expressed in the form

$$\omega_j=\sum_{1\leqq\nu\leqq n} f_{j\nu}\,dx_\nu \qquad (1\leqq j\leqq n),$$

where the $f_{j\nu}$ and the x_ν are rational functions on G defined over $\mathfrak{K}$ and defined at α, and $x_1,\cdots,x_n$ are algebraically independent over $\mathfrak{K}$, then the condition 2° above is equivalent to the condition that α satisfy a system of differential equations

$$\sum_{1\leqq\nu\leqq n} f_{j\nu}(\alpha)\,\delta_i x_\nu(\alpha)=a_{ij} \qquad (1\leqq i\leqq m,\ 1\leqq j\leqq n),$$

where the a_{ij} are elements of $\mathfrak{F}$.

By a *G-primitive extension* of $\mathfrak{F}$, we mean an extension of $\mathfrak{F}$ of the form $\mathfrak{F}\langle\alpha\rangle$, where α is a G-primitive over $\mathfrak{F}$.

It is not difficult to prove, making use of (3), that a necessary and sufficient condition for α to be a G-primitive over $\mathfrak{F}$ is that the field of constants of $\mathfrak{F}\langle\alpha\rangle$ be $\mathfrak{K}$ and $\sigma\alpha\cdot\alpha^{-1}\in G_{\mathfrak{K}}$ for every $\mathfrak{F}$-isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ onto an extension of $\mathfrak{F}$ in $\mathfrak{U}$. It follows almost immediately that every G-primitive extension of $\mathfrak{F}$ is a G-extension of $\mathfrak{F}$.

Suppose now that $\mathfrak{G}$ is a G-extension of $\mathfrak{F}$, so that there exists an injective rational homomorphism $\gamma\colon\mathfrak{G}\to G_{\mathfrak{K}}$, $\mathfrak{G}$ denoting the Galois group of $\mathfrak{G}$ over $\mathfrak{F}$. Regarding γ as a rational crossed homomorphism of $\mathfrak{G}$ into G, we find by the above that *if γ splits in $G_{\mathfrak{G}}$ (in particular if the cohomology set $H^1(\mathfrak{G},G)$ is trivial) then $\mathfrak{G}$ is a G-primitive extension of $\mathfrak{F}$.* It is known ([3] p. 887) that $H^1(\mathfrak{G},G)$ is trivial whenever G is the additive group of $\mathfrak{U}$, or the multiplicative group of $\mathfrak{U}$, or the general linear group $GL(r)$ for any r; this recaptures the characterizations ([2] p. 809 and [3] p. 891) of the strongly normal extensions corresponding to these groups as extensions of the form $\mathfrak{F}\langle\alpha\rangle$, where in the first case α is an element such that $\delta\alpha\in\mathfrak{F}$ for each derivation operator δ (extension by a *primitive* element), in the second case α is an element such that $\alpha^{-1}\delta\alpha\in\mathfrak{F}$ for each δ (extension by an *exponential* element), and in the third case α is an invertible $r\times r$ matrix such that $\alpha^{-1}\delta\alpha$ has its coordinates in $\mathfrak{F}$ for each δ (*Picard-Vessiot* extension).

If G is commutative, in particular if G is an abelian variety A, then we do not know in general that $H^1(\mathfrak{G},G)$, which is now a group, is trivial; but we do know ([3] p. 887) that every element of $H^1(\mathfrak{G},G)$ is of finite order,

so that for a suitable exponent $h > 0$ the cocycle γ^h splits in G_G. It follows in this case that there is a G-primitive α such that $\mathcal{F} \subset \mathcal{F}\langle\alpha\rangle \subset \mathcal{G}$ and the Galois group of $\mathcal{G}$ over $\mathcal{F}\langle\alpha\rangle$ equals the finite group $\mathfrak{G}(h)$ of all elements $\sigma \in \mathfrak{G}$ with $\sigma^h = 1$.

Returning to the general case of a G-extension $\mathcal{G}$ with arbitrary algebraic group G defined over $\mathcal{K}$, we see that thse normal chain (1) has a refinement

$$G \supset G^0 \supset H \supset H^0 \supset 1,$$

where H is the set of points $s \in G^0$ such that $s^h \in H$, and therefore the group variety $A = G^0/H$ is abelian and H/H^0 is finite abelian, and that the tower (2) has a corresponding refinement

$$\mathcal{F} \subset \mathcal{F}^0 \subset \mathcal{E} \subset \mathcal{E}^0 \subset \mathcal{G},$$

where $\mathcal{E}$ is A-primitive over $\mathcal{F}^0$ and $\mathcal{E}^0$ is an abelian extension of $\mathcal{E}$ of finite degree. Since H/H^0 is finite and H^0 is linear, it easily follows that H is birationally isomorphic with a linear algebraic group, and therefore that $\mathcal{G}$ is a Picard-Vessiot extension of $\mathcal{E}$. Thus *every strongly normal extension has a three-storey tower*

$$\mathcal{F} \subset \mathcal{F}^0 \subset \mathcal{E} \subset \mathcal{G}$$

in which $\mathcal{F}^0$ (the algebraic closure of $\mathcal{F}$ in $\mathcal{G}$) is a normal extension of $\mathcal{F}$ of finite degree, $\mathcal{E}$ is an A-primitive extension of $\mathcal{F}^0$ for some abelian variety A defined over $\mathcal{K}$, and $\mathcal{G}$ is a Picard-Vessiot extension of $\mathcal{E}$.

In the final § 3 the A-primitive extensions are given a function-theoretic characterization in the classical case in which $\mathcal{F}$ consists of functions of m complex variables meromorphic in some region of $\mathbf{C}^m$. It is shown that if A is an abelian variety of dimension n defined over the field $\mathbf{C}$ of complex numbers then *every A-primitive extension of $\mathcal{F}$ is of the form $\mathcal{F}(f_1(\zeta), \cdots, f_r(\zeta))$, where $f_1(z), \cdots, f_r(z)$ are abelian functions of $z = (z_1, \cdots, z_n)$ which generate a nondegenerate abelian function field, and $\zeta = (\zeta_1, \cdots, \zeta_n)$ is a family of n functions of m complex variables, each ζ_j being primitive over $\mathcal{F}$.*

1. Induced differentials. This section supplements the discussion of Rosenlicht [5]; a familiarity with the terminology and notation as well as with the methods and results of that paper is assumed. We find it convenient, for any element u of a vector space and any element v of the dual space, to denote the value of v at u by $\langle u, v\rangle$.

Let V be an algebraic variety defined over a field k, let K be a subfield of the universal domain $\mathcal{U}$ (not necessarily distinct from $\mathcal{U}$) with $k \subset K$, and let $\mathfrak{D}(K/k)$ denote the vector space over K formed by the derivations

of K over k. We recall (Rosenlicht [5], p. 54) that if α is a simple point on V which is rational over K then there is a unique K-linear mapping $\mathfrak{d} \rightarrow \mathfrak{d}_\alpha$, from $\mathfrak{D}(K/k)$ to the space of tangent vectors to V at α which are rational over K, defined by the condition that $\mathfrak{d}_\alpha f = \mathfrak{d} f(\alpha)$ $(f \in \mathfrak{o}_\alpha \cap k(V))$; also, if ω is any differential on V which is finite at α and defined over K, then the local component ω_α of ω at α is a K-linear mapping of the tangent space to V at α into $\mathcal{U}$ which maps the tangent vectors rational over K into K. The composite of the two mappings $\mathfrak{d} \rightarrow \mathfrak{d}_\alpha$ and ω_α then gives a K-linear mapping of $\mathfrak{D}(K/k)$ into K, that is, a differential of K over k; we call it the differential of K over k *induced by* ω *at* α, and following Rosenlicht denote it by $\omega(\alpha)$, or by $\omega_{K/k}(\alpha)$ when the greater detail is desirable.

If $f_i, g_i \in \mathfrak{o}_\alpha \cap k(V)$ $(1 \leqq i \leqq r)$ and $\omega = \sum f_i dg_i$ then

$$\langle \mathfrak{d}, \omega(\alpha) \rangle = \sum f_i(\alpha) \mathfrak{d} g_i(\alpha).$$

It is easy to see that if a subfield k_0 of k is also a field of definition of V and ω is defined over k_0 then $\langle \mathfrak{d}, \omega(\alpha) \rangle \in k_0(\alpha) \cdot k_0(\mathfrak{d} k_0(\alpha))$. Furthermore, if V' is also a variety defined over k, and $\phi \colon V \rightarrow V'$ is a rational mapping defined over k such that ϕ is defined at α and $\phi(\alpha)$ is a simple point on V', and ω' is a differential on V' finite at $\phi(\alpha)$ and defined over K, then $\omega'(\phi(\alpha)) = (\phi^* \omega')(\alpha)$.

We are interested in the case in which V is a group variety G and ω is an invariant (meaning left invariant) differential on G. Every point of G is simple, and ω is finite everywhere. In order to prove the result we require, we use the notion of a rational crossed homomorphism.

Let G and G' be connected algebraic groups. We say that G *operates on* G' *by* g if g is a group homomorphism of G into the group of automorphisms of G' such that the mapping $G \times G' \rightarrow G'$ defined by $(s, s') \rightarrow g(s)s'$ is rational. When such is the case $g(s)$ is a birational automorphism of G', defined over $k_0(s)$, whenever k_0 is a common field of definition of G, of G', and of the mapping $(s, s') \rightarrow g(s)s'$.

Suppose G operates on G' by g. A *rational crossed homomorphism* (or 1-*cocycle*) of G into G' (relative to g) is a rational mapping $\zeta \colon G \rightarrow G'$ such that, if k_0 is a common field of definition of G, of G', of the mapping $(s, s') \rightarrow g(s)s'$, and of ζ, and if (s, t) is a generic point of $G \times G$ over k_0, then $\zeta(st) = \zeta(s) \cdot g(s)\zeta(t)$. It is easy to see that this implies that ζ is defined at every point of G, $\zeta(st) = \zeta(s) \cdot g(s)\zeta(t)$ for all points s, t of G, and $\zeta(1) = 1$. (We denote the unity element of every multiplicative group by 1.) The notion of rational crossed homomorphism includes that of rational homomorphism, the latter corresponding to the case in which the operation g is trivial, that is $g(s)$ is the identity automorphism of G' for every $s \in G$.

Consider a rational crossed homomorphism $\zeta\colon G\to G'$. ζ induces a homomorphism $X\to\zeta X$ of the tangent space to G at 1 into the tangent space to G' at $1=\zeta(1)$; since these tangent spaces are canonically isomorphic (as vector spaces) to the respective Lie algebras $\mathfrak{L}(G)$ and $\mathfrak{L}(G')$ of G and G', we obtain a linear mapping $D\to\zeta D$ of $\mathfrak{L}(G)$ into $\mathfrak{L}(G')$. Thus, by definition, $(\zeta D)_1=\zeta D_1$ $(D\in L(G))$. (We use a point of a variety as a subscript to a derivation or differential on the variety to indicate the local component at that point.)

This being so for any rational crossed homomorphism, in particular it is so for any rational homomorphism, including the automorphism $g(s)\colon G'\to G'$.

Returning to ζ, we note that the crossed homomorphism property $\zeta(st)=\zeta(s)\cdot g(s)\zeta(t)$ can be written in the form $\zeta T_s=T_{\zeta(s)}g(s)\zeta$, where for any element v of any group we use T_v to denote the left multiplication mapping $x\to vx$ of that group into itself. It follows that for any $D\in\mathfrak{L}(G)$ we may write $\zeta D_s=\zeta T_sD_1=T_{\zeta(s)}g(s)\zeta D_1=T_{\zeta(s)}(g(s)\zeta D)_1=(g(s)\zeta D)_{\zeta(s)}$, so that

$$(g(s)\zeta D)_{\zeta(s)}=\zeta D_s \qquad (D\in\mathfrak{L}(G),s\in G).$$

Therefore the homomorphism $\omega'\to\zeta^*\omega'$ which ζ induces from the space of differentials on G' into the space of differentials on G has the property that $\langle D,\zeta^*\omega'\rangle(s)=\langle\zeta D_s,\omega'_{\zeta(s)}\rangle=\langle(g(s)\zeta D)_{\zeta(s)},\omega'_{\zeta(s)}\rangle=\langle g(s)\zeta D,\omega'\rangle(\zeta(s))$. But if ω' is an element of the space $\mathfrak{J}(G')$ of *invariant* differentials on G' then $\langle g(s)\zeta D,\omega'\rangle(t')$ is independent of the element $t'\in G'$, so that we may write

$$(4)\qquad \langle D,\zeta^*\omega'\rangle(s)=\langle g(s)\zeta D,\omega'\rangle(s)\ \ (D\in\mathfrak{L}(G),\omega'\in\mathfrak{J}(G'),s\in G).$$

In particular, a rational homomorphism $G\to G'$ induces vector space homomorphisms $\mathfrak{L}(G)\to\mathfrak{L}(G')$ and $\mathfrak{J}(G')\to\mathfrak{J}(G)$ each of which is the transpose of the other.

If $\nu\colon G\to G'$ is a rational homomorphism, and if $\zeta'\colon G'\to G''$ is a rational crossed homomorphism (G' operating on the connected algebraic group G'' by g'), then G operates on G'' by $g'\nu$, and $\zeta'\nu$ is a rational crossed homomorphism of G into G''; furthermore, $(\zeta'\nu)D=\zeta'(\nu D)$ $(D\in\mathfrak{L}(G))$.

We now apply these considerations to the differentials $\omega(\alpha)=\omega_{K/k}(\alpha)$ of K over k induced by an invariant differential ω on a connected algebraic group G; as before, G is defined over k, ω is defined over K, and $\alpha\in G_K$ (the group of points of G which are rational over K).

As examples of rational homomorphisms we have the two projections

$$\pi_j\colon G\times G\to G,\qquad \pi_j(s_1,s_2)=s_j,$$

the three injections

$$\iota_1: G \to G \times G, \quad \iota_1 s = (s, 1),$$
$$\iota_2: G \to G \times G, \quad \iota_2 s = (1, s),$$
$$\Delta: G \to G \times G, \quad \Delta s = (s, s),$$

the identity mapping $\iota: G \to G$, and the trivial endomorphism

$$\epsilon: G \to G, \qquad \epsilon s = 1.$$

Obviously $\pi_1\iota_1 = \pi_2\iota_2 = \pi_1\Delta = \pi_2\Delta = \iota$, and $\pi_1\iota_2 = \pi_2\iota_1 = \epsilon$; also $\epsilon D = 0$ $(D \in \mathfrak{L}(G))$. Furthermore, $\pi_1{}^*\mathcal{U}(G)$ is the subfield of $\mathcal{U}(G \times G)$ consisting of the rational functions f on $G \times G$ such that $f(s_1, s_2)$ is independent of s_2, and similarly for $\pi_2{}^*\mathcal{U}(G)$, so that $\mathcal{U}(G \times G)$ is precisely the compositum

$$\mathcal{U}(G \times G) = \pi_1{}^*\mathcal{U}(G) \cdot \pi_2{}^*\mathcal{U}(G). \tag{5}$$

For any $D \in \mathfrak{L}(G)$,

$$(\Delta D - \iota_1 D - \iota_2 D)_1 \pi_1{}^* f = (\pi_1(\Delta D - \iota_1 D - \iota_2 D)_1) f$$
$$= (\pi_1 \Delta D - \pi_1\iota_1 D - \pi_1\iota_2 D)_1 f = (\iota D - \iota D - \epsilon D)_1 f = 0$$

for all $f \in \mathfrak{o}_1$, so that

$$((\Delta D - \iota_1 D - \iota_2 D)\pi_1{}^* f)(s, t) = (\Delta D - \iota_1 D - \iota_2 D)_{(s,t)} \pi_1{}^* f = 0$$

provided $f \in \mathfrak{o}_s$; choosing (s, t) generic on $G \times G$ over a common field of definition of G, of D, and of f, we see that then $f \in \mathfrak{o}_s$ and

$$((\Delta D - \iota_1 D - \iota_2 D)\pi_1{}^* f)(s, t) = 0;$$

thus $(\Delta D - \iota_1 D - \iota_2 D)\pi_1{}^* f = 0$ for every $f \in \mathcal{U}(G)$. Similarly

$$(\Delta D - \iota_1 D - \iota_2 D)\pi_2{}^* f = 0$$

for every $f \in \mathcal{U}(G)$, so that by (5) $\Delta D = \iota_1 D + \iota_2 D$ $(D \in \mathfrak{L}(G))$.

Now, for each $t \in G$ let $\tau(t)$ denote the inner automorphism of G defined by t: $\tau(t)s = tst^{-1}$. Then G operates on itself by τ, and the reciprocation mapping

$$\rho: G \to G, \qquad \rho(s) = s^{-1}$$

is a rational crossed homomorphism. Also, $G \times G$ operates on G by $\tau\pi_2$, and the mapping

$$\chi: G \times G \to G, \qquad \chi(s, t) = st^{-1}$$

is a rational crossed homomorphism. Obviously $\chi\Delta = \epsilon$, $\chi\iota_1 = \iota$, and $\chi\iota_2 = \rho$. Therefore $0 = \epsilon D = \chi\Delta D = \chi\iota_1 D + \chi\iota_2 D = \iota D + \rho D$, so that

$$\rho D = -D \qquad (D \in \mathfrak{L}(G)).$$

It is easy to see that the two mappings $D \to \iota_j D$ of $\mathfrak{L}(G)$ into $\mathfrak{L}(G \times G)$ are injective. Also, if $D, E \in \mathfrak{L}(G)$ and $\iota_1 D + \iota_2 E = 0$ then $D = \iota D + \epsilon E = \pi_1 \iota_1 D + \pi_1 \iota_2 E = \pi_1(\iota_1 D + \iota_2 E) = 0$ and similarly $E = 0$. Therefore $\iota_1 \mathfrak{L}(G) \cap \iota_2 \mathfrak{L}(G) = \{0\}$, so that

$$\dim(\iota_1 \mathfrak{L}(G) + \iota_2 \mathfrak{L}(G)) = \dim \iota_1 \mathfrak{L}(G) + \dim \iota_2 \mathfrak{L}(G)$$
$$= 2 \dim \mathfrak{L}(G) = \dim \mathfrak{L}(G \times G),$$

whence

$$\mathfrak{L}(G \times G) = \iota_1 \mathfrak{L}(G) + \iota_2 \mathfrak{L}(G) \quad \text{(direct sum)}.$$

Since $\chi \iota_1 D - \pi_1 \iota_1 D + \pi_2 \iota_1 D = \iota D - \iota D + \epsilon D = 0$ and $\chi \iota_2 D - \pi_1 \iota_2 D + \pi_2 \iota_2 D = \rho D - \epsilon D + \iota D = 0$, we conclude that

$$\chi D = \pi_1 \hat{D} - \pi_2 \hat{D} \qquad (\hat{D} \in \mathfrak{L}(G \times G)). \tag{6}$$

For any $\alpha, \beta \in G_K$ (α, β) is a point of $(G \times G)_K$ so that for any $\mathfrak{d} \in \mathfrak{D}(K/k)$ we have the tangent vector $\mathfrak{d}_{(\alpha,\beta)}$ to $G \times G$ at (α, β) and $\mathfrak{d}_{(\alpha,\beta)}$ is rational over K; also, there is a unique $D \in \mathfrak{L}(G \times G)$ with $\hat{D}_{(\alpha,\beta)} = \mathfrak{d}_{(\alpha,\beta)}$. Since χ is obviously defined over k it follows that $\langle \mathfrak{d}, \omega(\alpha\beta^{-1}) \rangle = \langle \mathfrak{d}, \omega(\chi(\alpha, \beta)) \rangle = \langle \mathfrak{d}, (\chi^* \omega)(\alpha, \beta) \rangle = \langle \mathfrak{d}_{(\alpha,\beta)}, \chi^* \omega_{(\alpha,\beta)} \rangle = \langle D, \chi^* \omega \rangle (\alpha, \beta)$; by the general formula (4) this equals $\langle \tau(\beta) \chi \hat{D}, \omega \rangle (\alpha, \beta)$, and by (6) this in turn equals

$$\langle \tau(\beta)(\pi_1 \hat{D} - \pi_2 \hat{D}), \omega \rangle (\alpha, \beta) = \langle D, \pi_1{}^* \tau(\beta)^* \omega - \pi_2{}^* \tau(\beta)^* \omega \rangle (\alpha, \beta)$$
$$= \langle \mathfrak{d}_{(\alpha,\beta)}, (\pi_1{}^* \tau(\beta)^* \omega)_{(\alpha,\beta)} - (\pi_2{}^* \tau(\beta)^* \omega)_{(\alpha,\beta)} \rangle$$
$$= \langle \mathfrak{d}, (\pi_1{}^* \tau(\beta)^* \omega)(\alpha, \beta) - \pi_2{}^* \tau(\beta)^* \omega)(\alpha, \beta) \rangle$$
$$= \langle \mathfrak{d}, (\tau(\beta)^* \omega)(\alpha) - (\tau(\beta)^* \omega)(\beta).$$

Thus, $\omega(\alpha\beta^{-1}) = (\tau(\beta)^* \omega)(\alpha) - (\tau(\beta)^* \omega)(\beta)$. Applying this formula to $\tau(\beta^{-1})^* \omega$ instead of ω, and then replacing α by $\alpha\beta$, we obtain the desired formula (3) of the introduction.

2. G-extensions and G-primitives. In this section $\mathcal{F}$ denotes a commutative differential field of characteristic 0 with algebraically closed field of constants $\mathcal{C}$; the derivation operators of $\mathcal{F}$ are denoted by $\delta_1, \cdots, \delta_m$. $\mathcal{U}$ denotes a fixed universal extension of $\mathcal{F}$ (in the sense of [2] p. 768) and $\mathcal{K}$ denotes the field of constants of $\mathcal{U}$. We identify the derivation operators $\delta_1, \cdots, \delta_m$ in an obvious way with derivations of $\mathcal{U}$ over $\mathcal{K}$.

$\mathcal{U}$ is an algebraically closed extension of $\mathcal{F}$ of infinite transcendence degree, and therefore can be used as a universal domain for algebraic geometry. Whenever we introduce algebraico-geometric notions it always is in the algebraic geometry based on $\mathcal{U}$ as universal domain. The algebraic varieties we

have occasion to introduce, other than points, are actually defined over $\mathcal{B}$; if ω is a differential on such a variety, the symbol $\omega(\alpha)$ always denotes the differential $\omega_{\mathcal{U}/\mathcal{K}}(\alpha)$ of $\mathcal{U}$ over $\mathcal{K}$ induced by ω at α, so that α can be any simple point of the variety at which ω is finite (and therefore, in the case of an invariant differential on a group variety, can be any point whatsoever).

Let G be a connected algebraic group defined over $\mathcal{B}$; denote by $G_{\mathcal{K}}$ the group consisting of the points of G which are rational over $\mathcal{K}$.

It is apparent from the definition of induced differentials in § 1 that *if* $\gamma \in G$, *then a necessary and sufficient condition that* $\gamma \in G_{\mathcal{K}}$ *is that* $\langle \delta_i, \omega(\gamma) \rangle = 0$ *for every index* i *and every invariant differential* ω *on* G.

It follows from this that *if* $\alpha, \beta \in G$ *then a necessary and sufficient condition that* $\beta\alpha^{-1} \in G_{\mathcal{K}}$ *is that* $\langle \delta_i, \omega(\alpha) \rangle = \langle \delta_i, \omega(\beta) \rangle$ *for every index* i *and every invariant differential* ω *on* G. Indeed, setting $\gamma = \beta\alpha^{-1}$ we have, by formula (3),

$$\begin{aligned}\langle \delta_i, \omega(\beta) \rangle &= \langle \delta_i, \omega(\gamma\alpha) \rangle \\ &= \langle \delta_i, (\tau(\alpha^{-1})^*\omega)(\gamma) \rangle + \langle \delta_i, \omega(\alpha) \rangle,\end{aligned}$$

and as ω runs through the set of invariant differentials on G so does $\tau(\alpha^{-1})^*\omega$.

It is also obvious from the definition of induced differentials that *if* ω *is an invariant differential on* G *which is defined over* $\mathcal{B}$ *and* $\alpha \in G$, *then* $\langle \delta_i, \omega(\alpha) \rangle \in \mathcal{B}\langle \alpha \rangle$ $(1 \leqq i \leqq m)$. *If in addition* σ *is an isomorphism of the differential field* $\mathcal{B}\langle \alpha \rangle$ *over* $\mathcal{B}$ *onto an extension of* $\mathcal{B}$ *in* $\mathcal{U}$, *then* $\sigma\langle \delta_i, \omega(\alpha) \rangle = \langle \delta_i, \omega(\sigma\alpha) \rangle$ $(1 \leqq i \leqq m)$. To prove the second point we observe that there exist uniformizing coordinates $x_1, \cdots, x_n$ on G at α with each $x_j \in \mathcal{B}(G)$, and we may write $\omega = \sum f_j\, dx_j$ with each $f_j \in \mathfrak{o}_\alpha \cap \mathcal{B}(G)$; then

$$\langle \delta_i, \omega(\alpha) \rangle = \sum_{1 \leqq j \leqq n} f_j(\alpha)\, \delta_i x_j(\alpha),$$

and the result follows.

Let α be a G-primitive over $\mathcal{F}$. We recall from the introduction that this means: 1° the field of constants of $\mathcal{F}\langle \alpha \rangle$ is $\mathcal{B}$; 2° $\langle \delta_i, \omega(\alpha) \rangle \in \mathcal{F}$ for each index i and every invariant differential ω on G defined over $\mathcal{B}$. Then, for each isomorphism σ of $\mathcal{F}\langle \alpha \rangle$ over $\mathcal{F}$ onto an extension of $\mathcal{F}$ in $\mathcal{U}$, $\langle \delta_i, \omega(\sigma\alpha) \rangle = \sigma\langle \delta_i, \omega(\alpha) \rangle = \langle \delta_i, \omega(\alpha) \rangle$, so that the point $\gamma(\sigma) = \alpha^{-1} \cdot \sigma\alpha$ is in $G_{\mathcal{K}}$. Therefore $\mathcal{F}\langle \alpha \rangle \cdot \sigma(\mathcal{F}\langle \alpha \rangle) = \mathcal{F}\langle \alpha, \sigma\alpha \rangle = \mathcal{F}\langle \alpha, \gamma(\sigma) \rangle = \mathcal{F}\langle \sigma\alpha, \gamma(\sigma) \rangle$, so that $\mathcal{F}\langle \alpha \rangle$ is strongly normal over $\mathcal{F}$, and ([2] p. 768, cor. 5) the field of constants of $\mathcal{F}\langle \alpha \rangle \cdot \sigma(\mathcal{F}\langle \alpha \rangle)$ is $\mathcal{B}(\gamma(\sigma))$. Identifying the isomorphisms of $\mathcal{F}\langle \alpha \rangle$ over $\mathcal{F}$ with the unique automorphisms of $\mathcal{F}\langle \alpha \rangle \cdot \mathcal{K}$ over $\mathcal{F} \cdot \mathcal{K}$ to which they extend ([2], Ch. II, § 3), we find that $\alpha \cdot \gamma(\sigma\tau) = \sigma\tau\alpha = \sigma(\alpha \cdot \gamma(\tau)) = \alpha \cdot \gamma(\sigma) \cdot \gamma(\tau)$, so that $\gamma(\sigma\tau) = \gamma(\sigma)\gamma(\tau)$. Since $\gamma(\sigma) = 1$ only if $\sigma = 1$,

the mapping $\sigma \rightarrow \gamma(\sigma)$ is a birational isomorphism of the Galois group of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto an algebraic subgroup of $G_{\mathcal{K}}$. In particular, *every G-primitive extension of $\mathcal{F}$ is a G-extension of $\mathcal{F}$.*

Conversely, suppose that we are given a G-extension of $\mathcal{F}$, that is, a strongly normal extension $\mathcal{G}$ of $\mathcal{F}$ such that there exists an injective rational homomorphism $\gamma\colon \mathfrak{G} \rightarrow G_{\mathcal{K}}$, $\mathfrak{G}$ denoting the Galois group of $\mathcal{G}$ over $\mathcal{F}$.

If γ splits over $\mathcal{G}$, i.e. if there exists a point $\alpha \in G_{\mathcal{G}}$ such that $\gamma(\sigma) = \alpha \cdot \sigma\alpha^{-1}$ $(\sigma \in \mathfrak{G})$, then

$$\begin{aligned}\sigma\langle\delta_i, \omega(\alpha)\rangle &= \langle\delta_i, \omega(\sigma\alpha)\rangle = \langle\delta_i, \omega(\gamma(\sigma^{-1})\alpha)\rangle \\ &= \langle\delta_i, (\tau(\alpha^{-1})^*\omega)(\gamma(\sigma^{-1}))\rangle + \langle\delta_i, \omega(\alpha)\rangle = \langle\delta_i, \omega(\alpha)\rangle \quad (\sigma \in \mathfrak{G}),\end{aligned}$$

so that $\langle\delta_i, \omega(\alpha)\rangle \in \mathcal{F}$ for each index i and each invariant differential ω on G defined over $\mathcal{K}$, i.e. α is a G-primitive over $\mathcal{F}$. Since γ is injective, $\sigma\alpha = \alpha$ only if $\sigma = 1$, so that $\mathcal{F}\langle\alpha\rangle = \mathcal{G}$. Thus we have proved, in particular, that *if $\mathcal{G}$ is a G-extension of $\mathcal{F}$ with Galois group $\mathfrak{G}$ and if $H^1(\mathfrak{G}, G) = 1$ then $\mathcal{G}$ is a G-primitive extension of $\mathcal{F}$.*

If we do not assume that γ splits over $\mathcal{G}$, but suppose instead that G is commutative, then ([3] p. 887), for some integer $h > 0$, γ^h splits over $\mathcal{G}$: there exist a point $\alpha \in G_{\mathcal{G}}$ such that $\gamma(\sigma)^h = \alpha \cdot \sigma\alpha^{-1}$ $(\sigma \in \mathfrak{G})$. The same reasoning as above then proves: *If G is commutative and $\mathcal{G}$ is a G-extension of $\mathcal{F}$ with Galois group $\mathfrak{G}$, then there exist a G-primitive extension $\mathcal{H}$ of $\mathcal{F}$ and an integer $h > 0$ such that $\mathcal{F} \subset \mathcal{H} \subset \mathcal{G}$ and the Galois group of $\mathcal{G}$ over $\mathcal{H}$ is the finite group $\mathfrak{G}(h)$ of all elements σ of $\mathfrak{G}$ with $\sigma^h = 1$.*

3. Abelian extensions and abelian functions. In this section we consider the classical case in which $\mathcal{F}$ is a differential field consisting of functions meromorphic in some region of complex m-space $\boldsymbol{C}^m$, the derivation operators being the partial derivations $\partial/\partial x_1, \cdots, \partial/\partial x_m$ with respect to the m coordinate functions $x_1, \cdots, x_m$, and the field of constants of $\mathcal{F}$ being the complex number field $\boldsymbol{C}$. We seek to characterize the A-primitive extensions of $\mathcal{F}$ in function-theoretic terms, A denoting an arbitrary abelian variety defined over $\boldsymbol{C}$.

To this end we recall certain facts about abelian varieties defined over $\boldsymbol{C}$ and abelian functions. If Γ is a free abelian subgroup of $\boldsymbol{C}^n$ with $2n$ generators, the field $\mathcal{A}$ of all meromorphic functions on $\boldsymbol{C}^n$ admitting as periods the elements of Γ is called the abelian function field with period group Γ. $\mathcal{A}$ is said to be degenerate if by means of some invertible linear transformation on $\boldsymbol{C}^n$ the elements of $\mathcal{A}$ can all be expressed as meromorphic functions of fewer than n variables, that is, if there exist complex numbers

$k_1, \cdots, k_n$ not all 0 such that $\sum k_j \partial f/\partial z_j = 0$ $(f \in \mathcal{A})$, $z = (z_1, \cdots, z_n)$ denoting the usual coordinate functions on $\boldsymbol{C}^n$, and $\mathcal{A}$ is said to be nondegenerate otherwise. There are well known necessary and sufficient conditions on Γ for $\mathcal{A}$ to be nondegenerate (the period relations). When $\mathcal{A}$ is nondegenerate then Γ is of rank $2n$ over the field of real numbers, and there exist an abelian variety A of dimension n defined over $\boldsymbol{C}$ and a biholomorphic group isomorphism $\boldsymbol{C}^n/\Gamma \approx A_{\boldsymbol{C}}$ (these groups being endowed with their usual complex analytic structures). The composite of the canonical projection $\boldsymbol{C}^n \to \boldsymbol{C}^n/\Gamma$ and this isomorphism is a surjective holomorphic homomorphism $p\colon \boldsymbol{C}^n \to A_{\boldsymbol{C}}$ with kernel Γ such that $\mathcal{A} = \boldsymbol{C}(p(z))$. Conversely, if A is any abelian variety of dimension n defined over $\boldsymbol{C}$, there exist a nondegenerate abelian function field $\mathcal{A}$ and a surjective holomorphic homomorphism $p\colon \boldsymbol{C}^n \to A_{\boldsymbol{C}}$ exactly as above. These facts, or a reasonable facsimile of them, can be found in Chapter VI of Weil's book [6] (see especially his Theorem 5, page 130).

Now, $\mathcal{A}$ is obviously a differential field relative to the derivation operators $\partial/\partial z_1, \cdots, \partial/\partial z_n$, the field of constants of $\mathcal{A}$ being $\boldsymbol{C}$, and for each $k \in \boldsymbol{C}^n$ the translation mapping $f(z) \to f(z+k)$ $(f(z) \in \mathcal{A})$ is an automorphism of this differential field over $\boldsymbol{C}$. It follows by § 2 that if $(\omega_1, \cdots, \omega_n)$ denotes a basis of the space of invariant differentials on A defined over $\boldsymbol{C}$ then this automorphism maps $\langle \partial/\partial z_j, \omega_{j'}(p(z))\rangle$ onto $\langle \partial/\partial z_j, \omega_{j'}(p(z+k))\rangle$, which by § 2 equals $\langle \partial/\partial z_j, \omega_{j'}(p(z))\rangle + \langle \partial/\partial z_j, \omega_{j'}(p(k))\rangle = \langle \partial/\partial z_j, \omega_{j'}(p(z))\rangle$. Thus $\langle \partial/\partial z_j, \omega_{j'}(p(z))\rangle$ is invariant under every translation, that is, is a complex constant, which we denote by $a_{jj'}$ $(1 \leqq j \leqq n, 1 \leqq j' \leqq n)$. In other words, *$p(z)$ is an A-primitive over $\boldsymbol{C}$.* Because $\mathcal{A}$ is nondegenerate, the matrix $(a_{jj'})$ is invertible.

Shifting our attention to the differential field $\mathcal{F}$ of functions of m complex variables $x_1, \cdots, x_m$ meromorphic in some region D of $\boldsymbol{C}^m$ (m possibly different from n), let $\zeta(x) = (\zeta_1(x_1, \cdots, x_m), \cdots, \zeta_n(x_1, \cdots, x_m))$ be a sequence of n functions of $x = (x_1, \cdots, x_m)$ meromorphic in some subregion of D. Then

$$\begin{aligned}&\langle \partial/\partial x_i, \omega_{j'}(p(\zeta(x)))\rangle \\ &\qquad = \sum_j \partial\zeta_j(x)/\partial x_i \cdot \langle \partial/\partial z_j, \omega_{j'}(p(z))\rangle = \sum_j \partial\zeta_j(x)/\partial x_i \cdot a_{jj'},\end{aligned}$$

so that we may write the matrix equation

$$(\langle \partial/\partial x_i, \omega_{j'}(p(\zeta(x)))\rangle) = (\partial\zeta_j(x)/\partial x_i)(a_{jj'}). \tag{7}$$

Therefore *if each $\zeta_j(x)$ is primitive over $\mathcal{F}$, i.e. $\partial\zeta_j(x)/\partial x_i \in \mathcal{F}$ $(1 \leqq i \leqq m, 1 \leqq j \leqq n)$, then $p(\zeta(x))$ is an A-primitive over $\mathcal{F}$.*

Conversely, let q be any A-primitive over $\mathcal{F}$, so that we may write

$$\langle \partial/\partial x_i, \omega_{j'}(q)\rangle = f_{ij'}(x) \in \mathfrak{F} \qquad (1 \leqq i \leqq m, 1 \leqq j \leqq n).$$

It is easy to see that for each index j' the m functions $f_{ij'}(x)$ satisfy the integrability conditions $\partial f_{ij'}(x)/\partial x_{i'} = \partial f_{i'j'}(x)/\partial x_i$ $(1 \leqq i \leqq m, 1 \leqq l' \leqq m)$, and therefore for each j the m functions $g_{ij}(x)$ defined by the matrix equation $(g_{ij}(x)) = (f_{ij'}(x))(a_{jj'})^{-1}$ satisfy similar integrability conditions. Consequently there exist functions $\zeta_j(x)$ meromorphic in some subregion of D such that $\partial\zeta_j(x)/\partial x_i = g_{ij}(x)$ $(1 \leqq i \leqq m, 1 \leqq j \leqq n)$. Therefore by (7)

$$\begin{aligned}(\langle \partial/\partial x_i, \omega_{j'}(p(\zeta(x)))\rangle)& \\ &= (g_{ij}(x))(a_{jj'}) = (f_{ij'}(x)) = (\langle \partial/\partial x_i, \omega_{j'}(q)\rangle).\end{aligned}$$

It follows by § 2 that $q \cdot p(\zeta(x))^{-1} \in A_C$, i.e. $q \cdot p(\zeta(x))^{-1} = p(k)$ for some $k = (k_1, \cdots, k_n) \in \boldsymbol{C}^n$, whence $q = p(\zeta(x) + k)$. Since each $\zeta_j(x) + k_j$ is primitive over $\mathfrak{F}$, we have the following result: *In the classical case of a differential field $\mathfrak{F}$ of functions meromorphic in a region of complex m-space $\boldsymbol{C}^m$ with field of constants $\boldsymbol{C}$, a necessary and sufficient condition that a point $q \in A$ be an A-primitive over $\mathfrak{F}$ is that $q = p(\zeta)$, where $\zeta = (\zeta_1, \cdots, \zeta_n)$ and each ζ_j is primitive over $\mathfrak{F}$.* This means that the A-primitive extensions of $\mathfrak{F}$ are precisely the extensions of the form $\mathfrak{F}\langle p(\zeta)\rangle = \mathfrak{F}(p(\zeta))$, and therefore: *Every A-primitive extension of $\mathfrak{F}$ is obtained by adjoining to $\mathfrak{F}$ finitely many functions $f_1(\zeta), \cdots, f_r(\zeta)$,* where *$f_1(z), \cdots, f_r(z)$ are abelian functions which generate the nondegenerate abelian function field $\mathfrak{A}$ and $\zeta = (\zeta_1, \cdots, \zeta_n)$ is a sequence of n functions which are primitive over $\mathfrak{F}$.*

COLUMBIA UNIVERSITY.

REFERENCES.

[1] Iacopo Barsotti, "Un teorema di struttura per le varieta gruppali," *Rendiconti dell'Academia Nazionale dei Lincei*, vol. 18 (1955), pp. 43-50.

[2] E. R. Kolchin, "Galois theory of differential fields," *American Journal of Mathematics*, vol. 75 (1953), pp. 753-824.

[3] ———, "On the Galois theory of differential fields," *ibid.*, vol. 77 (1955), pp. 868-894.

[4] Maxwell Rosenlicht, "Some basic theorems on algebraic groups," *ibid.*, vol. 78 (1956), pp. 401-443.

[5] ———, "A note on derivations and differentials on algebraic varieties," *Portugaliae mathematica*, vol. 16 (1957), pp. 43-55.

[6] André Weil, *Introduction à l'étude des variétés kaehlériennes* (Actualités scientifiques et industrielles 1267), 1958.

THE ALGEBRAIC DETERMINATION OF THE GENUS OF KNOTS.*

By LEE NEUWIRTH.[1]

1. Introduction. The genus of a tame knot in S^3 was first defined by Seifert [1]. Although it was defined purely geometrically, he showed that the degree of the Alexander Polynomial, a purely algebraically defined invariant, is at most twice the genus. This inequality relating a geometric invariant to an algebraic invariant was sharpened by Crowell [2] and Murasugi [3] to an equality for alternating knots, which are defined geometrically.

In the course of investigating the commutator subgroup of knot groups I was able to prove a structure theorem which shows that *the genus of a knot is, under a certain simple algebraic condition, determined algebraically by the knot group*. Furthermore, this algebraic condition is often easily checked, and if satisfied, the genus is effectively computable.

Using a result of E. Rapaport [4] and the structure theorem mentioned above, equality between the degree of the Alexander Polynomial and twice the genus is established under an algebraic hypothesis quite different from the geometric hypothesis of Crowell and Murasugi.

A number of corollaries to the structure theorem will be proved, some of which relate to the center of a knot group, and constitute a start in the direction of proving the following:

Conjecture 1. A knot group G is the group of a torus knot if and only if G has a non-trivial center.

2. Hypothesis and notation. Throughout this paper k will denote an arbitrary but fixed non-trivial, polygonal knot in S^3.

By the group of the knot k, is meant $\pi_1(S^3 - k)$, which will be denoted G.

g will denote the genus of k.

Unless the base point is critical $\pi_1(X, x_0)$ will be denoted $\pi_1(X)$.

* Received November 11, 1959.

[1] The helpful suggestions of the referee are gratefully acknowledged. Part of this paper was prepared while the author held N. S. F. Post-doctoral Fellowship 49164.

The free product with amalgamation of two groups E, F on subgroups B, A respectively, will be denoted $E \underset{C}{*} F$, where $C \approx B \approx A$.

$\sim$ will denote "is homotopic to."

3. A construction. The infinite cyclic unbranched covering of $S^3 - k$ will be constructed in this section.

Let $S \subset S^3$ denote a polyhedral orientable surface of minimal genus (g) spanned by k [1].

Let X denote the unbranched covering of $S^3 - k$ corresponding to $[G, G]$.

Let U be an open regular neighborhood (in the sense of [5]) of the interior of S. Then $\dot{U} \supset k$, and $\bar{U}$ is a manifold with boundary [5]. Since S is orientable $\dot{U} - k$ has two polyhedral components, call them S_1 and S_2. Notice that genus of $S_1 =$ genus of $S_2 =$ genus of $S = g$.

X will be constructed by matching copies of $S^3 - (U \cup k)$ along S_1 and S_2. More precisely, let $X_0 = S^3 - (U \cup k)$. Let $\{X_i\}^{+\infty}_{i=-\infty}$ denote a countable collection of disjoint copies of X_0, where X_i may be identified with X_0 by a natural homeomorphism $f_i \colon X_i \to X_0$. In each X_i let ${}_iS_1$, ${}_iS_2$ denote $f_i^{-1}(S_1)$, $f_i^{-1}(S_2)$ respectively. Since U is a regular neighborhood of the interior of S, the interior of S is a deformation retract of $\bar{U} - k$ (see [5], p. 291); furthermore, this deformation retraction may be adjusted so as to map each S_j homeomorphically onto the interior of S; let g_j be this homeomorphism for $j = 1$ or 2.

Now match each X_i with X_{i-1} along ${}_{i-1}S_1$, ${}_iS_2$ as follows. Consider the space X' obtained from $\{X_i\}^{+\infty}_{i=-\infty}$ by factoring by the relation,

$$x = f^{-1}_{i+1} g_2^{-1} g_1 f_i(x) \quad \text{if } x \in {}_iS_1 \text{ for some } i.$$

This matches ${}_iS_1$ with ${}_{i+1}S_2$ for all i. X' may be pictured with the help of this diagram.

$$\begin{array}{r|cr|l} X_{i-1} & & X_i & X_{i+1} \\ & & & \\ {}_{i-1}S_1 & {}_iS_2 & {}_iS_1 & {}_{i+1}S_2. \end{array}$$

It must be shown that X' is homeomorphic to X. Recall $G/[G, G] \approx Z$. Clearly one sheet of X is homeomorphic to $S^3 - (U \cup k \cup S_1)$. The elements, t^n, of the group of covering translations of X thus correspond in a one to one manner to the maps $(f^{-1}_{i+1} f_i)^n$. Since X' covers $S^3 - k$ in the obvious way, and since X' has an infinite cyclic group of covering translations ($\{f^{-1}_{i+1} f_i\}_{i \in Z}$) it follows that X' is homeomorphic to X. It will be assumed from now on that the X_i are imbedded in X'.

Two lemmas are needed before we can proceed to calculate $\pi_1(X) = [G, G]$.

4. Two lemmas.

LEMMA 1. *The inclusion map* $h: S_1 \to \dot{U}$ *induces a monomorphism* $h^{\#}: \pi_1(S_1) \to \pi_1(\dot{U})$.

Proof. Since k is knotted, S_i is of genus greater than zero, hence no power of k, considered as an element of $\pi_1(S_i)$ is null homotopic. $\dot{U}$ is constructed by identifying $\bar{S}_1$ with $\bar{S}_2$ along k. A simple application of the Van Kampen Theorem thus gives

$$\pi_1(\dot{U}) \approx \pi_1(\bar{S}_1) \underset{\pi_1(k)}{*} \pi_1(\bar{S}_2).$$

Since the inclusion $S_1 \subset \bar{S}_1$ clearly induces an isomorphism of the fundamental group, $\pi_1(S_1)$ is imbedded monomorphically in $\pi_1(\dot{U})$ by a map induced by the inclusion h.

LEMMA 2. *The inclusion map* $i: S_1 \to (S^3 - U)$ *induces a monomorphism* $i^{\#}: \pi_1(S_1) \to \pi_1(S^3 - U)$.

Proof. Suppose the lemma is false, and that α is a closed curve on S_1, such that $\alpha \sim 0$ in $S^3 - U$, $\alpha \not\sim 0$ on S_1. According to Lemma 1, $\alpha \not\sim 0$ on $\dot{U}$. Since S_1 is polyhedral, α may be assumed polygonal so that $(S^3 - U)$ and α satisfy the hypothesis of the Loop Theorem ([6], Theorem 15.1 and Theorem 1), thus we may assume that α is simple. According to Dehn's Lemma [7], α bounds a non-singular polyhedral disc in $S^3 - U$. If we make a cut (Umschaltung) along this disc and its boundary we obtain a new surface S'_1 which is bounded by k. If the curve α separates S_1, then because $\alpha \not\sim 0$ on S_1, the new surface S'_1 has lower genus than $\bar{S}_1$, which contradicts the assumption that the genus of $\bar{S}_1$ is minimal. If α does not separate S_1, then compare the Euler characteristic of $\bar{S}_1$, $\chi(\bar{S}_1)$, with that of $\bar{S}'_1$. Since the cut adds one vertex, one edge, and two faces, $\chi(S'_1) - \chi(\bar{S}_1) = 2$, hence S'_1 has lower genus than $\bar{S}_1$, so we again arrive at a contradiction. The existence of the curve α thus leads to a contradiction, so the lemma is proved.

Remark. Lemmas 1 and 2 obviously remain valid if S_2 is substituted for S_1. Lemma 2 also remains valid if $S^3 - (U - k)$ is substituted for $S^3 - U$.

5. The structure theorem.

THEOREM 1. *If* $[G, G]$ *is finitely generated, it is free of rank* $2g$, *where* g *is the genus of* k.

If $[G, G]$ is not finitely generated, then either it is:

A) *a non-trivial free product with amalgamation on a free rank 2g, and may be written in the form*

$$\cdots \underset{F_{2g}}{*} A \underset{F_{2g}}{*} A \underset{F_{2g}}{*} A \underset{F_{2g}}{*} A \underset{F_{2g}}{*} A \underset{F_{2g}}{*} \cdots,$$

where F_{2g} is free of rank 2g, and the amalgamations are all proper and identical, or

B) *locally free, and a direct limit of free groups of rank 2g.*

Proof. By virtue of Lemma 2, the last remark in 4 and the identification of ${}_iS_1$ and ${}_{i+1}S_2$, the following diagram is valid for every i,

$$\pi_1(X_i) \xleftarrow[{}_if_1]{} \pi_1({}_iS_1) \xrightarrow[{}_{i+1}f_2]{} \pi_1(X_{i+1})$$

where the ${}_if_j$ are monomorphisms.

By a simple application of the Van Kampen theorem, the fundamental group of $X_i \cup X_{i+1}$ is the direct limit of the above system. This direct limit is a free product with amalgamated subgroup, $\pi_1(X_i) \underset{\pi_1(S)}{*} \pi_1(X_{i+1})$.

Let

$$Y_n = X_0 \cup X_1 \cup \cdots X_{n-1} \cup X_n, \qquad n \geqq 0$$

$$Y_{-n} = X_{-1} \cup X_{-2} \cup \cdots X_{-n+1} \cup X_{-n}, \qquad n \geqq 1$$

$$Y_\infty = \bigcup_{i=0}^{\infty} X_i \qquad Y_{-\infty} = \bigcup_{i=-1}^{-\infty} X_i.$$

Using the fact that each factor in a free product with amalgamation is contained as a subgroup in the free product with amalgation ([8], p. 32) and proceeding inductively it is clear that from the above diagram one obtains,

$$(1) \qquad \pi_1(Y_n) = \pi_1(Y_{n-1} \underset{\pi_1(S)}{*} \pi_1(X_n)$$
$$= (((\pi_1(X_0) \underset{\pi_1(S)}{*} \pi_1(X_1)) \underset{\pi_1(S)}{*} \pi_1(X_2)) \cdots \underset{\pi_1(S)}{*} \pi_1(X_{n-1})) \underset{\pi_1(S)}{*} \pi_1(X_n).$$

$$(2) \qquad \pi_1(Y_{-n}) = \pi_1(Y_{-n+1}) \underset{\pi_1(S)}{*} \pi_1(X_{-n})$$
$$= (((\pi_1(X_{-1}) \underset{\pi_1(S)}{*} \pi_1(X_{-2})) \underset{\pi_1(S)}{*} \pi_1(X_{-3})) \cdots \underset{\pi_1(S)}{*} \pi_1(X_{-n+1})) \underset{\pi_1(S)}{*} \pi_1(X_{-n}).$$

$$\pi_1(Y_\infty) = \varinjlim_{n \geqq 0} (Y_n, \phi_n),$$ where ϕ_n denotes the inclusion isomorphism of $\pi_1(Y_n)$ in $\pi_1(Y_{n+1})$ ([8], p. 32), and

$$\pi_1(Y_{-\infty}) = \varinjlim_{n \geqq 0} (Y_{-n}, \rho_{-n}),$$ where ρ_{-n} denotes the inclusion isomorphism of $\pi_1(Y_{-n+1})$ in $\pi_1(Y_{-n})$ ([8], p. 32).

It follows then that $\pi_1(Y_0) \subset \pi_1(Y_\infty)$, and $\pi_1(Y_{-1}) \subset \pi_1(Y_{-\infty})$. This fact and the diagram above imply that

$$\pi_1(X) = \pi_1(Y_{-\infty} \cup Y_\infty) = \pi_1(Y_{-\infty}) \underset{\pi_1(S)}{*} \pi_1(Y_\infty).$$

Note that if k is of genus g, then $\pi_1(S)$ is free of rank $2g$.

Suppose one of the maps, ${}_0f_i\colon \pi_1({}_0S_1) \to \pi_1(X_0)$ is not onto, say for $i = 1$. Then no ${}_jf_i$ will be onto, so that

$$\pi_1(Y_\infty) \subsetneq \pi_1\,(Y_{-1} \cup Y_\infty) \subsetneq \pi_1(Y_{-2} \cup Y_\infty) \cdots$$

and

$$\pi_1(X) = \bigcup_{n=1}^{\infty} \pi_1(Y_{-n} \cup Y_\infty)$$

so that $\pi_1(X)$ is not finitely generated. But if ${}_0f_i\colon \pi_1({}_0S_i) \to \pi_1(X_0)$ is onto for $i = 1, 2$, then all ${}_jf_i$ are onto so that

$$\pi_1(S) \approx \pi_1(X_0) \approx \pi_1(Y_1) \approx \pi_1(Y_n) \approx \pi_1(Y_\infty) \approx \pi_1(Y_{-1} \cup Y_\infty)$$
$$\approx \pi_1(Y_{-n} \cup Y_\infty) \approx \pi_1(Y_{-\infty} \cup Y_\infty) \approx \pi_1(X)$$

and $\pi_1(X)$ is free of rank $2g$. Hence if $[G, G] = \pi_1(X)$ is finitely generated ${}_0f_1$ and ${}_0f_2$ are onto and $\pi_1(X)$ is free of rank $2g$. This proves the first assertion of Theorem 1.

If neither of the mappings ${}_0f_i$ is onto, then $\pi_1(X) = \pi_1(Y_{-\infty}) \underset{\pi_1(S)}{*} \pi_1(Y_\infty)$ is a proper free product with amalgamation and may be written as $\varinjlim_{n>0} \pi_1(Y_n \cup Y_{-n})$, where each $\pi_1(Y_n \cup Y_{-n})$ is a proper free product with amalgamation of $\pi_1(Y_n)$ and $\pi_1(Y_{-n})$ on a free group of rank $2g$ isomorphic to $\pi_1(S)$, so by virtue of equations 1) and 2, A) is proven. Suppose one of the maps, ${}_0f_i$, is onto, say ${}_0f_2$, and the other, ${}_0f_1$, is not, then

$$\pi_1(Y_\infty \cup Y_{-n}) \subsetneq \pi_1(Y_\infty \cup Y_{-n-1})$$

and

$$\pi_1(S) \approx \pi_1(Y_0) \approx \pi_1(Y_1) \approx \pi_1(Y_n) \approx \pi_1(Y_\infty).$$

Since Y_∞ is homeomorphic to $Y_{-n} \cup Y_\infty$,[2] $\pi_1(Y_\infty) \approx \pi_1(Y_{-n} \cup Y_\infty)$, and hence each of the groups in the direct system

$$\pi_1(Y_\infty) \subsetneqq \pi_1(Y_\infty \cup Y_{-1}) \subsetneqq \pi_1(Y_\infty \cup Y_{-2}) \cdots$$

is free of rank $2g$. Of course $\pi_1(X) = \varinjlim_{n>0} \pi_1(Y_\infty \cup Y_{-n})$. Since the mappings in this direct system are all inclusions, it follows that any finitely generated subgroup $H \subset \pi_1(X)$ lies in some $\pi_1(Y_\infty \cup Y_{-n})$ and so H is a subgroup of a free group and hence free [9]; in other words, $\pi_1(X) = [G, G]$ is locally free, and B) is proven. This completes the proof of the theorem.

6. Application. In [1] H. Seifert showed (Satz 4) that the torus knot of type (p, q) has genus $\frac{1}{2}(p-1)(q-1)$. Theorem 1 allows an alternative proof to Seifert's result, since the rank of the commutator subgroup of $G = (a, b : a^p = b^q)$ may be easily shown to be $(p-1)(q-1)$.

7. Corollaries. In this section a number of corollaries to Theorem 1 are given, many of which concern the center of a knot group.

COROLLARY 1. *If $[G, G]$ is finitely generated, then no orientable surface of minimal genus spanned by k is algebraically knotted.*[3]

Proof. If $[G, G]$ is finitely generated, $\pi_1(S) \approx \pi_1(X_0) \approx \pi_1(S^3 - S)$, hence $\pi_1(S^3 - S)$ is free.

COROLLARY 2. *The center of $[G, G]$ is trivial.*

Proof. If $[G, G]$ is finitely generated, it is free of rank > 1 according to Theorem 1, hence the corollary follows.

If $[G, G]$ is not finitely generated, it is either a free product with amalgamation on a centerless group and hence centerless according to [8, p. 32], or else locally free and non-abelian and hence centerless. Q. E. D.

COROLLARY 3. *If k is non-trivial, G contains a free group of rank n for any $n \leqq \infty$.*

Proof. $G \supset [G, G] \supset \pi_1(S)$ according to the proof of Theorem 1, and $\pi_1(S)$ is free of rank $2g$ ($g > 0$ since k is knotted). Since the free group of rank 2 contains a free group of countable rank, the corollary follows.

---

[2] The homeomorphism may be taken to be the covering translation t^n restricted to $Y_{-n} \cup Y_\infty$.

[3] A surface $S \subset S^3$ is algebraicallyknotted if $\pi_1(S^3 - S)$ is not free.

COROLLARY 4. *The comutator series of a non-cyclic knot group does not terminate after a finite number of steps.*

Proof. If the commutator series of a knot group terminated after a finite number of steps, the same would be true for any subgroup (see for example [8], p. 179). By Corollary 3 there exists a subgroup with a non-finite commutator series ([8], p. 36), hence the corollary is proved.

COROLLARY 5. *The center, Z, of a knot group, G, is cyclic.*

Proof. Corollary 2 implies $Z \cap [G, G] = 1$, hence

$$G/[G, G] \supset Z/Z \cap [G, G] = Z.$$

Since $G/[G, G]$ is cyclic, the corollary is proved.

Remark. Of course if Z is non-trivial, Z is infinite cyclic.

The existence of a non-trivial center in a knot group has a rather interesting consequence.

THEOREM 2. *If a knot group, G, has a non-trivial center, Z, then Z intersects every abelian subgroup of rank 2 non-trivially.*

Proof. According to Corollary 5, $Z = (z:\ )$. Suppose

$$G \supset H = (x:\ ) \oplus (y:\ ),$$

then the subgroup HZ is free abelian, and of rank 2 according to [10].

Let s and t generate HZ; then the following equations are valid for some integers m_i, n_i;

$$n_1 s + m_1 t = z,$$

$$n_2 s + m_2 t_2 = x,$$

$$n_3 s + m_3 t = y.$$

Since x and y generate H, and H is not cyclic, $n_3 x - n_2 y \neq 0$, $m_3 x - m_2 y \neq 0$; thus

$$(m_3 n_2 - m_2 n_3)(s) = m_3 x - m_2 y \neq 0,$$

$$(n_3 m_2 - n_2 m_3)(t) = n_3 x - n_2 y \neq 0.$$

Upon multiplying each of these equations by n_1 and m_1 respectively, we deduce

$$0 \neq K_1 z = K_2 x + K_3 y$$

so that $Z \cap H \neq 1$, which was to be proved.

Since $[G, G] \cap Z = 1$ by Corollary 2, Theorem 2 implies

COROLLARY 6. *If a knot group G has a non-trivial center, then $[G, G]$ has no abelian subgroups of rank 2.*

I have shown in [11] that *the group of any alternating knot which is not a torus knot has no center.* On the other hand it is easy to see that the group of any torus knot, $((a, b : a^m = b^n), (m, n) = 1)$ has an infinite cyclic center (generated by b^n). Thus Conjecture 1 is true for the class of alternating knots.

COROLLARY 7. *If $[G, G]$ is finitely generated, then the degree of the Alexander Polynomial is twice the genus of k.*

Proof. In [4] E. Rapaport has shown that when $[G, G]$ is free of rank $2g < \infty$, the degree of the Alexander polynomial is $2g$. This and Theorem 1 imply the corollary.

REFERENCES.

[1] H. Seifert, "Über das Geschlect von Knoten," *Mathematische Annalen*, vol. 110 (1935), pp. 571-592.

[2] R. H. Crowell, "Genus of alternating link types," *Annals of Mathematics*, vol. 69 (1959), pp. 258-275.

[3] K. Murasugi, "On the genus of the alternating knot, I, II," *Journal of the Mathematical Society of Japan*, vol. 10 (1958), pp. 94-105, pp. 235-248.

[4] E. Rapaport, "On the commutator subgroup of a knot group," *Annals of Mathematics*, vol. 71 (1960), pp. 157-162.

[5] J. H. C. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society*, vol. 45 (1939), pp. 243-327.

[6] C. D. Papakyriakopoulos, "On solid tori," *ibid.*, (3), vol. 7 (1957), pp. 281-299.

[7] ———, "On Dehn's lemma and the asphericity of knots," *Annals of Mathematics*, vol. 66 (1957), pp. 1-26.

[8] K. A. Kurosh, *The Theory of Groups, Vol. II*, New York, 1955.

[9] O. Schreien, "Die Untergruppen der freien Gruppen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 5 (1926), pp. 161-183.

[10] P. Conner, "On the action of a finite group on $S^n \times S^n$," *Annals of Mathematics*, vol. 66 (1957), pp. 586-588.

[11] L. Neuwirth, *Knot Groups*, Princeton University Doctoral Thesis, June, 1959.

AN ULTRAHYPERBOLIC EQUATION WITH AN INTEGRAL CONDITION.*[1]

By O. G. OWENS.

Introduction. The ultrahyperbolic differential equation with four independent variables,

$$\partial^2 u/\partial t_1^2 + \partial^2 u/\partial t_2^2 = \partial^2 u/\partial x_1^2 + \partial^2 u/\partial x_2^2, \tag{0.1}$$

is of special importance in the investigation of the Hilbert problem of determining in 3-dimensional cartesian space all metrics for which the geodesics are straight lines. G. Hamel [1] showed its relationship to the Hilbert proposal and established for (0.1) a characteristic value problem, the data being analytic in some of the variables. F. John [2] has treated with elegance other characteristic value problems for the same equation, his data not being restricted by analyticity requirements. Furthermore, N. S. Piskunov [3] has treated the conical characteristic value problem for (0.1).

In 1923 J. Hadamard [4] asserted that the ultrahyperbolic equation does not seem to possess any "properly posed" limit problem; that is, an auxiliary limit condition which when taken in conjunction with the differential equation assures both the uniqueness and the existence of the solution. In 1946 I. G. Petrowsky [5] made a similar statement: "There is a sizable class of partial differential equations for which we do not know of any correctly posed limit problem. The so-called ultrahyperbolic equation seems to be one of these."

It seems to the author that the Hadamard-Petrowsky statement applies primarily to non-characteristic value problems and but one such problem, which seems to be a simple natural limit problem for (0.1), will be discussed in this paper. The problem will be concerned with solutions of (0.1) which are defined on the 4-dimensional cartesian region consisting of all points (x_1, x_2, t_1, t_2) where (t_1, t_2) is an arbitrary point of the cartesian (t_1, t_2)-space and (x_1, x_2) any point of a given bounded region $G(X)$ of (x_1, x_2)-space. These solutions are assumed to vanish for all (t_1, t_2) whenever (x_1, x_2) is at

* Received December 13, 1959.

[1] This work was supported by the National Science Foundation, Contract No. G6093.

the boundary $G^*(X)$ of $G(X)$. Moreover, the auxiliary limit condition required of the solution is that

$$(0.2) \qquad \int_0^\infty u\,ds = F(x_1, x_2, \theta) \qquad (t_1 + it_2 = se^{i\theta}),$$

where the prescribed function $F(x_1, x_2, \theta)$ is of period 2π in θ, regular on its domain, and vanishes with sufficient smoothness at the boundary $G^*(X)$. Subject to these restrictions it is shown that there exists a unique solution of $(0,1)$, Theorem 3.1 and Theorem 12.1, and that the integral $(0,2)$ is only conditional convergent, Theorem 12.2. An evident conclusion from these theorems and their proofs will be the existence of a bi-unique mapping of the above class of solutions of (0.1) onto denumerable sequences of harmonic functions defined on the unit circle.

1. Notation. $x = (x_1, x_2)$ and $t = (t_1, t_2)$ will denote any two points of the cartesian (x_1, x_2)-space and (t_1, t_2)-space, and $u = u(x, t) = u(x_1, x_2, t_1, t_2)$ any real-valued function of x and t. $G(X)$ will indicate any simply connected bounded region of (x_1, x_2)-space with boundary, $G^*(X)$, consisting of a simple closed curve formed from a finite number of analytic arcs. $G(T)$ signifies the entire (t_1, t_2)-space and $G = G(X) \times G(T)$ the 4-dimensional region consisting of all points $(x, t) = (x_1, x_2, t_1, t_2)$ with $x \in G(X)$ and $t \in G(T)$. The closure of any set E will be denoted by $\bar{E}$. The euclidean elements of area on $G(X)$ and $G(T)$ are abbreviated to $dx = dx_1 dx_2$ and $dt = dt_1 dt_2$, and the related double integrals are written as

$$\int_{G(X)} u\,dx = \int\int_{G(X)} u\,dx_1 dx_2, \qquad \int_{G(T)} u\,dt = \int\int_{G(T)} u\,dt_1 dt_2.$$

It is convenient to indicate the repeated Laplacian by

$$\Delta_x{}^l u = \Delta_x(\Delta_x{}^{l-1} u) \text{ or } \Delta_t{}^l u = \Delta_t(\Delta_t{}^{l-1} u) \qquad (l = 0, 1, \cdots),$$

where

$$\Delta_x u = \partial^2 u/\partial x_1{}^2 + \partial^2 u/\partial x_2{}^2, \qquad \Delta_t u = \partial^2 u/\partial t_1{}^2 + \partial^2 u/\partial t_2{}^2,$$

and $D_x{}^l u(D_t{}^l u)$ will stand for any partial derivative of u of order l $(D_x{}^0 u = D_t{}^0 u = u)$ with respect to the variables $(x_1, x_2)((t_1, t_2))$. Furthermore, the index in any sum will always vary successively over the positive integers. Finally, the notation "Theorem 12.2" refers to the second theorem of Section twelve.

2. Characteristic functions with estimates. The characteristic functions and values corresponding to the Laplacian operator and the region $G(X)$

are indicated by $\phi_n = \phi_n(x)$ and λ_n $(n = 1, 2, \cdots)$. They satisfy the following equations:

$$\Delta_x \phi_n + \lambda_n \phi_n = 0 \qquad x \in G(X), \tag{2.1}$$

$$\phi_n(x) = 0 \qquad x \in G^*(X), \tag{2.2}$$

$$\int_{G(X)} \phi_n^2 \, dx = 1. \tag{2.3}$$

In addition, the dependence of the maxima of the moduli of ϕ_n and its partial derivatives upon λ_n are known. Indeed, A. Hammerstein [6] has shown the existence of a positive constant C which is independent of x and n and for which

$$| D_x^k \phi_n | \leqq C \lambda_n^{1+k} \qquad (k = 0, 1, \text{ and } n = 1, 2, \cdots), \tag{2.4}$$

provided that $G^*(X)$ is formed from a finite number of analytic arcs. Furthermore, J. Schauder [7] has investigated the general linear second order elliptic differential equation and has determined bounds for the second and third order partial derivatives of the solution. It follows, as an easy consequence of the bounds of Schauder and those of Hammerstein, that

$$| D_x^k \phi_n | \leqq C \lambda_n^{1+k} \qquad (k = 2, 3, \text{ and } n = 1, 2, \cdots). \tag{2.5}$$

3. Statement of existence theorem.

THEOREM 3.1. *There exists a $u(x,t)$, with $D_x^l u$ and $D_t^l u$ $(l = 0, 1, 2)$ continuous on $\bar{G}(X) \times G(T)$, which simultaneously satisfies the ultrahyperbolic equation*

$$\Delta_t u = \Delta_x u \qquad (x, t) \in G, \tag{3.1}$$

the boundary condition

$$u(x, t) = 0 \qquad (x, t) \in G^*(X) \times G(T), \tag{3.2}$$

and the integral condition

$$\int_0^\infty u(x, t) \, ds = F(x, \theta) \qquad (t_1 + it_2 = se^{i\theta}), \tag{3.3}$$

the integral converging uniformly in (x, θ).

It is assumed that the arbitrarily prescribed $F(x, \theta)$ is of period 2π in θ and in addition possesses the following properties: (i) $\Delta_x^l F$ $(l = 0, 1, \cdots, 5)$ and $D_x^1(\Delta_x^l F)$ $(l = 0, 1, \cdots, 4)$ are continuous on

$$\bar{G}(X) \times I, \qquad I = \{\theta \mid 0 \leqq \theta \leqq 2\pi\}.$$

(ii) $\Delta_x{}^l F$ $(l=0,1,\cdots,5)$ vanishes at $G^*(X)$. (iii) the normal derivative $\partial\Delta_x F/\partial\nu$ $(l=0,1,\cdots,5)$ is bounded on $G^*(X)$. (iii) $F(x,\theta)$ and $\Delta_x{}^2F$ satisfy a Hölder condition with respect to θ, uniformly in (x,θ), with Hölder exponent $\alpha>\frac{1}{2}$; that is, there is a constant H independent of x and θ such that

$$
(3.4)\qquad \begin{aligned} &|F(x,\theta_2)-F(x,\theta_1)|\leqq H\,|\theta_2-\theta_1|^\alpha,\\ &|\Delta_x{}^2F(x,\theta_2)-\Delta_x{}^2F(x,\theta_1)|\leqq H\,|\theta_2-\theta_1|^\alpha \end{aligned}
$$

for $(x,\theta_i)\in G(X)\times I$, $i=1,2$.

4. Formal series representation of solutions. Assume that $u(x,t)$ is a solution of (0.1) which vanishes for $(x,t)\in G^*(X)\times G(T)$ and that $D_x{}^k u$ $(k=0,1,2)$ is continuous for $(x,t)\in\bar{G}(X)\times G(T)$. Then, as is well known (see [8], p. 369), $u(x,t)$ admits the series representation

$$
(4.1)\qquad u(x,t)=\sum\phi_n(x)\int_{G(X)}u(x,t)\phi_n(x)\,dx=\sum v_n(t)\phi_n(x),
$$

where

$$
(4.2)\qquad v_n(t)=\int_{G(X)}u(x,t)\phi_n(x)\,dx.
$$

Now, suppose that $u(x,t)$ fulfills (3.1)-(3.3) and $D_t{}^k u$ $(k=0,1,2)$ is continuous on $\bar{G}(X)\times G(T)$. Then, because of (4.2),

$$
(4.3)\qquad \begin{aligned} \Delta_t v_n &= \int_{G(X)}\phi_n\Delta_x u\,dx\\ &= \int_{G(X)}u\Delta_x\phi_n\,dx+\int_{G^*(X)}[\phi_n(\partial u/\partial\nu)-u(\partial\phi_n/\partial\nu)]\,dG^*; \end{aligned}
$$

where $\partial u/\partial\nu$ and $\partial\phi_n/\partial\nu$ denote the normal derivatives of u and ϕ_n at $G^*(X)$. Since $\partial\phi_n/\partial\nu$ $(n=1,2,\cdots)$ are uniformly bounded at $G^*(X)$, see (2.4) and (2.5), the line integral appearing in (4.3) vanishes. Thus $v_n(t)$ must be a solution of the reduced wave equation

$$
(4.4)\qquad \Delta_t v_n+\lambda_n v_n=0 \qquad t\in G(t).
$$

Besides, by reason of (3.3) and (4.2), $v_n(t)$ will also satisfy the integral condition

$$
(4.5)\qquad \int_0^\infty v_n(t)\,ds=\int_{G(X)}F(x,\theta)\phi_n\,dx\equiv F_n(\theta).
$$

Hence $v_n(t)$ simultaneously satisfies (4.4) and (4.5). Conversely, by Theorem 5.1, the conditions (4.4) and (4.5) assure the existence of a unique $v_n(t)$. Consequently, this completes the demonstration that the series (4.1)

represents the solution of the indicated problem provided that the series and its derivatives can be properly majorized and provided that (4.1) uniformly satisfies the integral condition (3.3).

5. The reduced wave equation. The following result was recently established in [9].

THEOREM 5.1. *There exists a unique* $v(t)$, *with* $D_t^l v$ $(l=0,1,2)$ *continuous on* $G(T)$, *which simultaneously satisfies the reduced wave equation*

$$\Delta_t v + \lambda v = 0 \qquad t \in G(T), \tag{5.1}$$

λ *a positive constant, and the integral condition*

$$\int_0^\infty v(t)\,ds = f(\theta) \qquad (t_1 + it_2 = se^{i\theta}), \tag{5.2}$$

the integral converging uniformly in θ.

It is assumed that the arbitrarily prescribed $f(\theta)$ is of period 2π in θ and in addition it posesses the following property: $f(\theta)$ fulfills a uniform Hölder condition with Hölder exponent $\alpha > \frac{1}{2}$; that is, there is a constant H independent of θ such that

$$|f(\theta_2) - f(\theta_1)| \leqq H\,|\theta_2 - \theta_1|^\alpha \tag{5.3}$$

for $\theta_i \in I$, $i = 1, 2$.

Finally, the solution $v(t)$ has the following explicit representation:

$$\begin{aligned} 2\pi^2 v(t) &= \pi\lambda^{\frac{1}{2}} \int_0^{2\pi} f(\phi) K_1(\lambda^{\frac{1}{2}}s, \phi - \theta)\,d\theta \\ &\quad + \lambda^{\frac{1}{2}} \int_0^{2\pi} d\phi f(\phi) \int_0^{\pi/2} (\sin\tau)^{-1} [K_2(\lambda^{\frac{1}{2}}s, \tau + \phi - \theta) + K_2(\lambda^{\frac{1}{2}}s, \tau - \phi + \theta)]\,d\tau, \end{aligned} \tag{5.4}$$

where

$$K_1(s,\theta) = \cos[s\sin\theta], \qquad K_2(s,\theta) = \sin[s\sin\theta]. \tag{5.5}$$

By Theorem 5.1 the function v_n satisfying (4.4) and (4.5) is uniquely and explicitly given by the formula (5.4) with λ replaced by λ_n and $f(\phi)$ by $F_n(\phi)$. The fact that $f(\phi) \equiv F_n(\phi)$ satisfies the Hölder condition (5.3) is shown in the proof of Lemma 11.1, and is a simple consequence of fact that $F(x,\theta)$ satisfies the Hölder condition (3.4). Hence, the formal series for $u(x,t)$ is

$$u(x,t) = \sum v_n(t)\phi_n(x), \tag{5.6}$$

where

$$
\begin{aligned}
2\pi^2 v_n(t) &= \pi\lambda_n^{\frac{1}{2}} \int_0^{2\pi} F_n(\phi) K_1(\lambda_n^{\frac{1}{2}}s, \phi-\theta)\, d\phi \\
&\quad + \lambda_n^{\frac{1}{2}} \int_0^{2\pi} d\phi F_n(\phi) \int_0^{\pi/2} (\sin\tau)^{-1} \\
&\quad \times [K_2(\lambda_n^{\frac{1}{2}}s, \tau+\phi-\theta) + K_2(\lambda_n^{\frac{1}{2}}s, \tau-\phi+\theta)]\, d\tau
\end{aligned} \tag{5.7}
$$

and

$$
F_n(\theta) = \int_{G(X)} F(x,\theta)\phi_n(x)\, dx. \tag{5.8}
$$

6. A bound for F_n.

LEMMA 6.1. *Assume that $\partial\Delta_x{}^l F/\partial\nu$ $(l=0,\cdots,k-1)$ is bounded on $G^*(X)$ and $\Delta_x{}^l F(x,\theta)$ $(l=0,\cdots,k-1)$ vanishes at $G^*(X)$. Then*

$$
F_n(\theta) = (-1)^k \lambda_n^{-k} \int_{G(X)} \phi_n \Delta_x{}^k F\, dx. \tag{6.1}
$$

Proof. The relation (6.1) follows by $k-1$ applications of Green's formula.

COROLLARY 6.1. *Retaining the hypotheses of Lemma* 6.1,

$$
|F_n(\theta)| \leqq B_k\lambda_n^{-k}, \text{ where } B_k{}^2 = \operatorname*{Max}_{\theta\in I} \int_{G(X)} [\Delta_x{}^k F]^2\, dx. \tag{6.2}
$$

Proof. Applying Schwarz's inequality to (6.1) and making use of (2.3) yields (6.2).

7. A bound for v_n.

LEMMA 7.1. *Set $\mu = \operatorname*{Max}_{n=1,2,\cdots} \lambda_n^{-\frac{1}{2}}$. Then, under the hypotheses of Lemma* 6.1,

$$
|v_n(t)| \leqq B_k\lambda_n^{1-k}(s+\mu). \tag{7.1}
$$

Proof. Define $v_1 = v_1(t,n)$ and $v_2 = v_2(t,n)$ by the following two equations:

$$
2\pi\lambda_n^{-\frac{1}{2}} v_1 = \int_0^{2\pi} F_n(\phi) K_1(\lambda_n^{\frac{1}{2}}s, \phi-\theta)\, d\phi, \tag{7.2}
$$

$$
\begin{aligned}
2\pi\lambda_n^{-\frac{1}{2}} v_2 = \pi^{-1} \int_0^{2\pi} d\phi F_n(\phi) \int_0^{\pi/2} (\sin\tau)^{-1} [K_2(\lambda_n^{\frac{1}{2}}s, \tau+\phi-\theta) \\
+ K_2(\lambda_n^{\frac{1}{2}}s, \tau-\phi+\theta)]\, d\tau.
\end{aligned} \tag{7.3}
$$

Therefore,

$$
v_n(t) = v_1(t,n) + v_2(t,n); \tag{7.4}
$$

so, an estimate for v_n will be obtained as soon as estimates for v_1 and v_2 are known. A bound for v_1 is immediately found by reason of (6.2) and (7.2), namely,

$$|v_1(t,n)| \leqq \lambda_n^{\frac{1}{2}} \operatorname{Max} |F_n(\phi)| \leqq B_k \lambda_n^{\frac{1}{2}-k}. \tag{7.5}$$

An estimate for v_2 will now be derived. Set $\delta = \phi - \theta$ and

$$I(s,\delta) = \int_0^{\pi/2} (\sin \tau)^{-1} [K_2(s, \tau + \delta) - K_2(s, \delta - \tau)] d\tau. \tag{7.6}$$

Hence, because of (6.2), (7.3) and (7.6),

$$\pi |v_2(t,n)| \leqq B_k \lambda_n^{\frac{1}{2}-k} \operatorname{Max} |I(\lambda_n^{\frac{1}{2}} s, \delta)|. \tag{7.7}$$

Moreover, because of the mean value theorem,

$$\begin{aligned} |K_2(s, \tau + \delta) - K_2(s, \delta - \tau)| &\leqq s |\sin(\tau + \delta) - \sin(\delta - \tau)| \\ &= 2s |\sin \tau \cos \delta| \end{aligned} \tag{7.8}$$

which implies

$$|I(s,\delta)| \leqq 2s \int_0^{\pi/2} d\tau \leqq \pi s. \tag{7.9}$$

It thus will follow from (7.7) and (7.9) that

$$|v_2(t,n)| \leqq s B_k \lambda_n^{1-k}. \tag{7.10}$$

Consequently, the bound (7.1) for v_n follows from the inequalities (7.5) and (7.10).

8. Bounds for the derivatives of v_1. By defining

$$\alpha_2 + i\alpha_1 = -\exp[-i\phi] \tag{8.1}$$

and recalling that $t_1 + it_2 = s \exp[i\theta]$, one finds that

$$K_1(\lambda_n^{\frac{1}{2}} s, \phi - \theta) = \cos[\lambda_n^{\frac{1}{2}}(\alpha_1 t_1 + \alpha_2 t_2)]. \tag{8.2}$$

Hence, using (7.2),

$$v_1 = \frac{1}{2\pi} \int_0^{2\pi} \lambda_n^{\frac{1}{2}} F_n(\phi) \cos[\lambda_n^{\frac{1}{2}}(\alpha_1 t_1 + \alpha_2 t_2)] d\phi \tag{8.3}$$

which implies that

$$\partial v_1 / \partial t_i = -\frac{1}{2\pi} \int_0^{2\pi} \alpha_i \lambda_n F_n(\phi) \sin[\lambda_n^{\frac{1}{2}}(\alpha_1 t_1 + \alpha_2 t_2)] d\phi \tag{8.4}$$

and

$$\partial^2 v_1 / \partial t_i \partial t_j = -\frac{1}{2\pi} \int_0^{2\pi} \alpha_i \alpha_j \lambda_n^{\frac{3}{2}} F_n(\phi) \cos[\lambda_n^{\frac{1}{2}}(\alpha_1 t_1 + \alpha_2 t_2)] d\phi. \tag{8.5}$$

From the last two equations, together with (6.2), one deduces the following estimates:

$$|\partial v_1/\partial t_i| \leqq B_k \lambda_n^{1-k} \tag{8.6}$$

and

$$|\partial^2 v_1/\partial t_i \partial t_j| \leqq B_k \lambda_n^{\frac{3}{2}-k}. \tag{8.7}$$

9. Bounds for the derivatives of v_2. By defining

$$\beta_2 + i\beta_1 = -\exp[-i(\tau+\phi)], \qquad \gamma_2 + i\gamma_1 = \exp[i(\tau-\phi)] \tag{9.1}$$

and placing

$$\Omega(t,\phi,\tau) \equiv K_2(\lambda_n^{\frac{1}{2}}s, \tau+\phi-\theta) + K_2(\lambda_n^{\frac{1}{2}}s, \tau-\phi+\theta), \tag{9.2}$$

one finds that

$$\Omega = \sin[\lambda_n^{\frac{1}{2}}(\beta_1 t_1 + \beta_2 t_2)] + \sin[\lambda_n^{\frac{1}{2}}(\gamma_1 t_1 + \gamma_2 t_2)]. \tag{9.3}$$

Hence, by (7.3),

$$v_2 = \frac{1}{2\pi^2}\int_0^{2\pi} d\phi\, \lambda_n^{\frac{1}{2}} F_n(\phi) \int_0^{\pi/2} \Omega(t,\phi,\tau)(\sin\tau)^{-1}\, d\tau; \tag{9.4}$$

and, by (9.3),

$$\begin{aligned} \partial\Omega/\partial t_i = \lambda_n^{\frac{1}{2}}\beta_i \cos[\lambda_n^{\frac{1}{2}}(\beta_1 t_1 + \beta_2 t_2)] \\ + \lambda_n^{\frac{1}{2}}\gamma_i \cos[\lambda_n^{\frac{1}{2}}(\gamma_1 t_1 + \gamma_2 t_2)], \end{aligned} \tag{9.5}$$

$$\begin{aligned} \partial^2\Omega/\partial t_i \partial t_j = -\lambda_n \beta_i \beta_j \sin[\lambda_n^{\frac{1}{2}}(\beta_1 t_1 + \beta_2 t_2)] \\ - \lambda_n \gamma_i \gamma_j \sin[\lambda_n^{\frac{1}{2}}(\gamma_1 t_1 + \gamma_2 t_2)]. \end{aligned} \tag{9.6}$$

From (9.5) and (9.6) one can derive estimates for the derivatives with respect to τ of $\partial\Omega/\partial t_i$ and $\partial^2\Omega/\partial t_i\partial t_j$. Because of these estimates and the fact that the left-members of (9.5) and (9.6) vanish for $\tau = 0$, it follows from the mean value theorem that

$$|\partial\Omega/\partial t_i| \leqq 2\lambda_n \tau(s+\mu) \tag{9.7}$$

and

$$|\partial^2\Omega/\partial t_i \partial t_j| \leqq 2\lambda_n^{\frac{3}{2}}(s+2\mu), \tag{9.8}$$

where $\mu = \underset{n=1,2,\cdots}{\operatorname{Max}} \lambda_n^{-\frac{1}{2}}$.

It is now possible to find estimates for the first and second partial derivatives of $v_2(t,n)$. First, these derivatives are given by the formulas

$$\partial v_2/\partial t_i = \frac{1}{2\pi^2}\int_0^{2\pi} d\phi \lambda_n^{\frac{1}{2}} F_n(\phi) \int_0^{\pi/2} (\sin\tau)^{-1}(\partial\Omega/\partial t_i)\, d\tau \tag{9.9}$$

and

$$(9.10) \qquad \partial^2 v_2/\partial t_i \partial t_j = \tfrac{1}{2\pi^2} \int_0^{2\pi} d\phi \lambda_n^{\frac{1}{2}} F_n(\phi) \int_0^{\pi/2} (\sin \tau)^{-1} (\partial \Omega/\partial t_i \partial t_j)\, d\tau.$$

Thus, by (6.2), (9.7) and (9.9),

$$(9.11) \quad |\, \partial v_2/\partial t_i \,| \leqq 2(s+\mu) B_k \lambda_n^{\frac{3}{2}-k} \int_0^{\pi/2} \tau (\sin \tau)^{-1} d\tau \leqq 2(s+\mu) C B_k \lambda_n^{\frac{3}{2}-k}$$

and, by (6.2), (9.8) and (9.10),

$$(9.12) \quad |\, \partial^2 v_2/\partial t_i \partial t_j \,| \leqq 2(s+\mu) B_k \lambda_n^{2-k} \int_0^{\pi/2} \tau (\sin \tau)^{-1} d\tau \leqq 2(s+\mu) C B_k \lambda_n^{2-k},$$

with a suitably chosen constant C.

As $v_n(t) = v_1(t,n) + v_2(t,n)$, the inequalities (8.6) and (9.11) yield

$$(9.13) \qquad |\, \partial v_n/\partial t_i \,| \leqq B_k \lambda_n^{\frac{3}{2}-k} [\mu + 2C(s+\mu)],$$

while (8.7) and (9.12) yield

$$(9.14) \qquad |\, \partial^2 v_n/\partial t_i \partial t_j \,| \leqq B_k \lambda_n^{2-k} [\mu + 2C(s+\mu)].$$

10. Majorization of series. The solution $u(x,t)$ has been defined by the formal series

$$(10.1) \qquad u(x,t) = \sum v_n(t) \phi_n(x),$$

where $v_n(t)$ is uniquely determined by means of the conditions (4.4) and (4.5). The series (10.1), together with its formal partial derivatives, can be majorized by making use of the estimates given in Sections two, seven and nine. In particular, the bounds (2.4), (7.1) and (9.13) give the following majorizations:

$$(10.2) \qquad \sum |\, v_n(t) \phi_n(x) \,| \leqq C B_k (s+\mu) \sum \lambda_n^{2-k},$$

$$(10.3) \qquad \sum |\, v_n D_x^1 \phi_n \,| \leqq C B_k (s+\mu) \sum \lambda_n^{3-k}$$

$$(10.4) \qquad \sum |\, D_t^1 v_n \phi_n \,| \leqq C B_k (\mu + 2C(s+\mu)) \sum \lambda_n^{\frac{5}{2}-k},$$

while the bounds (2.4), (2.5), (7.1) and (9.14) give

$$(10.5) \qquad \sum |\, v_n D_x^2 \phi_n \,| \leqq C B_k (s+\mu) \sum \lambda_n^{4-k}$$

$$(10.6) \qquad \sum |\, D_t^2 v_n \phi_n \,| \leqq C B_k (\mu + 2C(s+\mu)) \sum \lambda_n^{3-k}.$$

By reason of the well-known inequality, $\sum \lambda_n^{-2} < \infty$ (see [8], p. 130), and the above estimates, the following lemma and corollary have been established.

LEMMA 10.1. *The series* (10.2)-(10.6) *are uniformly convergent, for* $k=6$, *on the closure of* $G(X)\times G_N(T)$, *where* N *is any positive real number and* $G_N(T)=\{(t_1,t_2)\mid -N\leqq t_1,t_2\leqq N\}$.

COROLLARY 10.1. $u(x,t)$ *of* (10.1) *and the partial derivatives* $D_t{}^l u$, $D_s{}^l u$ $(l=1,2)$ *exist and are continuous on the closure of* $G(X)\times G_N(T)$. *Furthermore,* $u(x,t)$ *is a solution of the ultrahyperbolic equation* (0.1).

By reason of Corollary 10.1, the proof of Theorem 3.1 has now been completed except for the fact that the integral condition (3.3) has still to be established.

11. Verification of integral condition. Here in this section it will be shown that $u(x,t)$ is such that

$$\int_0^\infty u(x,t)\,ds=F(x,\theta), \tag{11.1}$$

the integral converging uniformly with respect to $(x,\theta)\in G(X)\times I$. Since the series (10.1) is uniformly convergent on $G(X)\times G_N(T)$,

$$\int_0^N u(x,t)\,ds=\sum\phi_n(x)\int_0^N v_n(t)\,ds. \tag{11.2}$$

In order to discuss the passage to the limit $(N\to\infty)$ in (11.2), it is necessary to use the following lemma.

LEMMA 11.1. *The integral*

$$\int_0^\infty v_n(t)\,ds \qquad (n=1,2,\cdots) \tag{11.3}$$

converges uniformly with respect to (n,θ).

Proof. It was proved in the paper [9] that the integral (11.3) converges uniformly in θ when n is fixed. That the convergence is uniform with respect to (n,θ) will now be established. First, the integrals (3.7) and (3.8) of paper [9], page 394, converge to zero uniformly with respect to (n,θ) provided that the Hölder coefficients A_n of F_n are uniformly bounded and the Hölder exponents have a greatest lower bound which is greater than $\frac{1}{2}$. But this is clear from (5.8), by virtue of Schwarz's inequality, the normalization condition (2.3), and the Hölder condition, (3.4), satisfied by F. Furthermore, because of the assertion, the integral (4.3) of [9], p. 394, converges uniformly in (n,θ). Consequently, the condition (3.4) of the present paper is sufficient to assure that the integrals (11.3) converge uniformly in (n,θ).

It is now possible to investigate the passage to the limit, $N \to \infty$, in equation (11.2).

LEMMA 11.2. *The* $u(x,t)$ *of* (10.1) *is such that the integral relation* (11.1) *holds uniformly in* (x,θ).

Proof. Since $F(x,\theta)$ is subject to the conditions of Theorem 3.1, it can be expanded into the series

$$F(x,\theta) = \sum F_n(\theta)\phi_n(x), \tag{11.4}$$

where $F_n(\theta)$ is defined by equation (5.8). Thus,

$$\int_0^N u(x,t)\,ds - F(x,\theta) = \sum\Big[\int_0^N v_n(t)\,ds - F_n(\theta)\Big]\phi_n(x). \tag{11.5}$$

Now, setting

$$F_{nk}(\theta) \equiv \int_{G(X)} \phi_n(x)\Delta_x{}^k F(x,\theta)\,dx, \tag{11.6}$$

the Lemma 6.1 asserts that

$$F_n(\theta) = (-1)^k \lambda_n{}^{-k} F_{nk}(\theta), \tag{11.7}$$

with k to be presently fixed. Assuming that $\Delta_x{}^k(x,\theta)$ is such that

$$|\Delta_x{}^k F(x,\theta_2) - \Delta_x{}^k F(x,\theta_1)| < H\,|\theta_2-\theta_1|^\alpha \qquad (\alpha > \tfrac{1}{2}), \tag{11.8}$$

H an absolute constant, it follows that $F_{nk}(\theta)$ will satisfy the same type of Hölder condition with Hölder coefficient uniformly bounded. Hence, if $v_{nk} = (-1)^k \lambda_n{}^k v_n$, then $\Delta_t v_{nk} + \lambda_n v_{nk} = 0$ and

$$\int_0^\infty v_{nk}(t)\,ds = F_{nk}(\theta), \tag{11.9}$$

the integral converging uniformly in (n,θ). Thus, the equation (11.5) can be put in the form

$$\int_0^N u(x,t)ds - F(x,\theta) = \sum(-1)^k\lambda_n{}^{-k}\Big[\int_0^N v_{nk}(t)ds - F_{nk}(\theta)\Big]\phi_n(x). \tag{11.10}$$

Furthermore, given an $\epsilon > 0$, the uniform convergence of (11.9) implies the existence of an $N_0(\epsilon)$, independent of (n,θ), for which

$$\Big|\int_0^N v_{nk}(t)\,ds - F_{nk}(\theta)\Big| < \epsilon \tag{11.11}$$

for all $N \geqq N_0(\epsilon)$. Consequently, (11.10) and (11.11) yield the next inequality.

$$
(11.12) \qquad \begin{aligned} &\left| \int_0^N u(x,t)\,ds - F(x,\theta) \right| \\ &\qquad \leqq \epsilon \sum \lambda_n^{-k} \,|\, \phi_n(x) \,| \leqq \epsilon [\sum \lambda_n^{2(1-k)}]^{\frac{1}{2}} \cdot [\sum \phi_n^2(x) \lambda_n^{-2}]^{\frac{1}{2}}. \end{aligned}
$$

As the factor of ϵ in the last term in uniformly bounded for $k = 2$ (see [8], p. 130) and as (11.8) holds by hypothesis (see Theorem 3.1) for $k = 2$, the inequality (11.2) implies the validity of (11.1).

12. Uniqueness of solution and conditional convergence of integral.

THEOREM 12.1. *Assume there exists a* $u(x,t)$, *with* $D_x^l u$ *and* $D_t^l u$ $(l = 0, 1, 2)$ *continuous on* $\bar{G}(X) \times G(T)$, *which simultaneously satisfies the ultrahyperbolic equation*

$$
(12.1) \qquad \Delta_t u = \Delta_x u \qquad (x,t) \in G,
$$

the boundary condition

$$
(12.2) \qquad u(x,t) = 0 \qquad (x,t) \in G^*(X) \times G(T),
$$

and the integral condition

$$
(12.3) \qquad \int_0^\infty u(x,t)\,ds = 0,
$$

the integral converging uniformly in (x,θ). *Then* $u(x,t)$ *is necessarily identically zero.*

Proof. Because of the boundary condition and the differentiability conditions imposed on $u(x,t)$, it follows that it has the series representation (4.1) where $v_n(t)$ is given by (4.2). Therefore, by the uniformity of (12.3),

$$
(12.4) \qquad \int_0^\infty v_n(t)\,ds = \int_{G(X)} dx\, \phi_n(x) \int_0^\infty u(x,t)\,ds = 0.
$$

By Theorem 5.1 the only solution of (4.4) subject to (12.4) is $v_n(t) \equiv 0$ $(n = 1, 2, \cdots)$. Hence, because of the series representation (4.1), $u(x,t)$ is identically zero on G. That is, the solution of Theorem 3.1 is unique.

The fact that the integral (3.3) is only conditionally convergent is the context of the next theorem.

THEOREM 12.2. *Assume that* $u(x,t)$ *satisfies all of the hypotheses of Theorem* 12.1 *except that* (12.3) *is replaced by the requirement that the integral*

$$
(12.5) \qquad \int_0^\infty |\, u(x,t) \,|\, ds
$$

be uniformly convergent with respect to $(x,\theta) \in G(X) \times I$. *Then* $u(x,t)$ *is necessarily identically zero.*

Proof. On account of (4.2)

$$|v_n(t)| \leqq \int_{G(X)} |u(x,t)| \, |\phi_n(x)| \, dx, \tag{12.6}$$

and this implies that

$$\int_0^N |v_n(t)| \, ds \leqq \int_{G(X)} dx \, |\phi_n(x)| \int_0^N |u(x,t)| \, ds. \tag{12.7}$$

Thus, because of the uniform convergence of (12.5),

$$\int_0^\infty |v_n(t)| \, dt \leqq \int_{G(X)} dx \, |\phi_n(x)| \int_0^\infty |u(x,t)| \, ds < \infty, \tag{12.8}$$

the convergence being uniform in θ. Hence, by Corollary 3 of [9], p. 390, $v_n(t) \equiv 0$ $(n = 1, 2, \cdots)$ on $G(T)$. Consequently, because of the series representation (4.1), $u(x,t)$ is identically zero on G.

WAYNE STATE UNIVERSITY.

REFERENCES.

[1] G. Hamel, "Über die Geometrieen in denen die Geraden die Kürzesten sind," *Mathematischen Annalen*, vol. 57 (1903), pp. 231-264.

[2] F. John, "The ultrahyperbolic differential equation with four independent variables," *Duke Mathematical Journal*, vol. 4 (1938), pp. 300-322.

[3] N. S. Piskunov, "On the characteristic problem for equations of ultrahyperbolic type," *Doklady Akademie Nauk SSSR*, vol. 59 (1948), pp. 439-442.

[4] J. Hadamard, *Lectures on Cauchy's Problem*, Yale University Press, 1923, pp. 4 and 39.

[5] I. G. Petrowsky, "On some problems of the theory of partial differential equations," *American Mathematical Society Translations*, Number 12.

[6] A. Hammerstein, "Über Entwicklung gegebener Funktionen nach Eigenfunktionen von Randwertaufgaben," *Mathematische Zeitschrift*, vol. 27 (1927), pp. 269-311, especially pp. 290-294.

[7] J. Schauder, "Über den Zusammenhang zwischen der Eindeutigkeit und Lösbarkeit partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Mathematische Annalen*, vol. 106 (1932), pp. 661-721, especially pp. 687-688.

[8] R. Courant and D. Hilbert, Methods of mathematical physics, vol. 1, New York, 1953.

[9] O. G. Owens, "A characterization of the everywhere regular solution of the reduced wave equation," *Transactions of the American Mathematical Society*, vol. 88 (1958), pp. 388-399.

QUADRATIC FORMS OVER FIELDS OF CHARACTERISTIC 2.*

By Chih-Han Sah.

In [13], Minkowski solved the problem of rational equivalence for quadratic forms. This was extended by Hasse to algebraic number fields, [10], [11]. Witt [16], and Arf [1], then developed general theories of quadratic forms over fields of characteristic $\neq 2$ and $=2$ respectively. These works together solved the field equivalence problem of quadratic forms over sufficiently "nice" fields, e.g. global and local fields.

The problem of integral equivalence for quadratic forms over a global field remains open. In a sequence of papers, O'Meara [14], [15], solved the problem of integral equivalence for quadratic forms over a local field of characteristic $\neq 2$. Earlier, Arf [2], solved the same problem for binary and ternary quadratic forms over local fields of characteristic 2. The purpose of the present investigation, which formed the basic part of the author's thesis, is to give a solution to the integral equivalence problem for quadratic forms in any finite number of variables over a local field of characteristic 2. The main results of this paper are contained in Theorem 4.6, for modular lattices, and in Theorem 5.5 for lattices in general.

Familiarity of [1] is assumed. Basic notions on quadratic forms may be found in [4], [5] and [9]. Basic results on fields and algebras may be found in [3], [5] and [17].

I wish to thank Prof. O'Meara for introducing me to the theory of quadratic forms in his seminar given during the spring of 1958 as well as the many interesting conversations on quadratic forms during the preparation of this paper.

1. Notations. Throughout this paper, Ω shall denote a finite field of characteristic 2. $k=\Omega\langle\langle\pi\rangle\rangle$, the formal power-series field over Ω with uniformizer π, $\pi^\infty=0$. $\mathfrak{o}=\Omega[[\pi]]$, the ring of integral power series. $\Omega[\pi^{-1}]$, the polynomial ring in π^{-1} over Ω. 0 and λ are fixed representatives of $\Omega/\mathfrak{P}\Omega$, where $\mathfrak{P}x=x^2+x$. ord is the ordinal function in k and Ord is the ordinal function in $k/\mathfrak{P}k$ defined by $\mathrm{Ord}(\Delta)=\max\{\mathrm{ord}(x)\mid x\in\Delta\}$, where $\Delta\in k/\mathfrak{P}k$.

* Received November 24, 1959.

By a quadratic space (respectively lattice) U over k, we mean a vector space (respectively free module) of finite type over k (respectively $\mathfrak{o}$), together with a map $Q\colon U \to k$ such that $Q(ax) = a^2Q(x)$ for $a \in k$ (respectively $a \in \mathfrak{o}$) and $x \in U$, and such that $\langle x, y\rangle = Q(x+y) + Q(x) + Q(y)$ is bilinear. If U is the direct sum of U_1 and U_2 and $\langle U_1, U_2\rangle = 0$, then we shall write $U = U_1 \oplus U_2$; such decompositions will be called orthogonal. $\dim U =$ rank of U over k (or $\mathfrak{o}$). $R(U) = \{x \in U \mid Q(x) = 0, \langle x, U\rangle = 0\}$ and $K(U) = \{x \in U \mid \langle x, U\rangle = 0\}$ will be called the radical and core of U respectively. It is easily seen that $R(U)$ and $K(U)$ are orthogonal summands of U, and that the structure of U is uniquely determined by $R(U)$ and the natural quadratic structure on $U/R(U)$. Thus, we shall always assume that the given quadratic spaces or quadratic lattices are non-degenerate, i. e. $R(U) = 0$. U is called non-defective when $K(U) = 0$. It is easy to see that $\dim K(U) \leqq [k\colon k^2] = 2$.

If M is a quadratic lattice, then $k \otimes_{\mathfrak{o}} M$ receives a natural quadratic structure, this quadratic space shall be denoted by kM. If $a \neq 0 \in k$, then the space (or lattice) M together with the quadratic map $a \circ Q$ defined by $(a \circ Q)(x) = aQ(x)$ for $x \in M$, shall be denoted by $a \circ M$. This operation is called scaling by a. Homomorphisms which preserve the quadratic structures will be called representations. Isomorphic representations are called isometries and denoted by $\cong$. From the non-degeneracy hypothesis, it follows easily that representations are always monomorphic.

The quadratic lattice $\mathfrak{o}x$ such that $Q(x) = a \neq 0 \in k$ will be denoted by (a). The quadratic lattice $\mathfrak{o}x + \mathfrak{o}y$ such that $Q(x) = a$, $\langle x, y\rangle = b \neq 0$ and $Q(y) = c$, where $a, b, c \in k$, shall be denoted by $\begin{pmatrix} & b & \\ a & & c\end{pmatrix}$. The subscript k will be used to denote the quadratic spaces spanned by the corresponding lattices. If the basis element x is to be emphasized, we shall write $\mathfrak{o}x = (a)$, otherwise, we shall use $\mathfrak{o}x \cong (a)$, similarly for binary and higher dimensional lattices and spaces.

If V is a quadratic space, then we shall let $C(V)$ denote the Clifford algebra of V, cf. [5; Chap. 9], [9]. If V is non-defective, then $C(V)$ is a k-central simple algebra and $C((1)_k \oplus V)/\mathfrak{R} \approx C(V)$, where $\mathfrak{R}$ is the radical of the algebra $C((1)_k \oplus V)$, cf. [1; p. 151]. In general, $C(U \oplus V)$ is isomorphic to $C(U) \oplus_k C(V)$. If $V = \begin{pmatrix} & 1 & \\ a & & c\end{pmatrix}_k$, then the class of $C(V)$ in the Brauer group of k is denoted by $[a, c]$. It can be shown, [17; p. 135] that, $[a, c] = [c, a]$, $[ad^2, c] = [a, cd^2]$, $[d^2, c] = 1$ and $[a, b + c] = [a, b][a, c]$, where the group operation in the Brauer group is written multiplicatively. Using these, it is easy to see that $[a, c] = 1$ or $[\pi^{-1}, \pi\lambda]$ $(\neq 1)$.

If $V \cong \oplus \begin{pmatrix} & b_i \\ a_i & & c_i \end{pmatrix}_k$ is a non-defective quadratic space, then the Arf invariant is denoted by $\Delta(V) = \sum a_i c_i b_i^{-2} + \mathfrak{P}k \in k/\mathfrak{P}k$, cf. [1], [6], [12] or [18]. We shall need the following fact:

LEMMA 1.1. *Let $\Delta \in k/\mathfrak{P}k$ and $a \in k$, then,*

1). $\mathrm{Ord}(\Delta) > 0$ *implies that* $\Delta = \mathfrak{P}k$ *and* $\mathrm{Ord}(\Delta) = \infty$;

2). $\mathrm{Ord}(\Delta) = 0$ *implies that* $\Delta = \lambda + \mathfrak{P}k$ *and conversely*;

3). $\mathrm{Ord}(\Delta) < 0$ *implies that* $\mathrm{Ord}(\Delta) = 2s + 1$, *$s$ an integer*;

4). $a \in \Delta$ *and* $\mathrm{ord}(a) = 2s + 1 < 0$ *implies that* $\mathrm{Ord}(\Delta) = 2s + 1$;

5). $a = \pi^{-1}b^2 + D + \mathfrak{P}c$, *where* $b \in \Omega[\pi^{-1}]$ *and* $D = 0$ *or* λ; b, D *and* $\mathfrak{P}c$ *are uniquely determined by* a. *In particular, if* $\mathrm{ord}(a) = 2s + 1 < 0$, *then* $\mathrm{ord}(\pi^{-1}b^2 + D) = \mathrm{ord}(\pi^{-1}b^2) = 2s + 1 < \mathrm{ord}(\mathfrak{P}c)$; *if* $\mathrm{ord}(a) = 2s < 0$, *then* $\mathrm{ord}(\mathfrak{P}c) = 2s < \mathrm{ord}(\pi^{-1}b^2 + D)$; *if* $\mathrm{ord}(a) = 0$, *then* $b = 0$; *and, if* $\mathrm{ord}(a) > 0$, *then* $b = D = 0$.

Proof. Applying Hensel's Lemma [7; p. 43] to $X^2 + X + d$, where $d \in \mathfrak{o}$, we get 1) and 2) immediately.

3). Let $a \in \Delta$ such that $\mathrm{ord}(a) = \mathrm{Ord}(\Delta) < 0$. Suppose that $\mathrm{ord}(a) = 2s$ and $a = \pi^{2s}(b^2 + \pi c^2)$, where $s < 0$, $\mathrm{ord}(b) = 0$ and $\mathrm{ord}(c) \geqq 0$. Then $a' = a + \mathfrak{P}(\pi^s b) \in \Delta$ and $\mathrm{ord}(a') > 2s$, a contradiction.

4). Let $a' \in \Delta$ such that $\mathrm{ord}(a) = 2s + 1 < \mathrm{ord}(a')$. Then, $a + a' \in \mathfrak{P}k$ and $\mathrm{ord}(a + a') = 2s + 1 < 0$. Thus, $a + a' = b^2 + b$ and $\mathrm{ord}(b) < 0$. Therefore, $\mathrm{ord}(b^2 + b) = 2 \cdot \mathrm{ord}(b) = 2s + 1$, a contradiction.

5) follows easily from the first four results. Q. E. D.

Let L be a quadratic lattice. $s(L) = \{\langle x, y\rangle \mid x, y \in L\}$ is called the scale of L. $Q(L) = \{Q(x) \mid x \in L\}$ is called the norm of L. $q(L) = \{Q(x) + s(L) \mid x \in L\}$ is called the norm group of L. For $-\infty < i \leqq \infty$, $L(i) = \{x \in L \mid \langle x, L\rangle \subseteq \pi^i\mathfrak{o}\}$ is called the i-th invariant sublattice of L and $q_i(L) = \{Q(x) + \pi^i\mathfrak{o} \mid x \in L(i)\}$ is called the i-th norm group of L.

It is easily seen that the scale is an $\mathfrak{o}$-module of finite type in k and that the norm groups are $\mathfrak{o}^2$-modules of finite type in k. If A, B are subsets of k, then let $A + B = \{a + b \mid a \in A, b \in B\}$. If $L = L_1 \oplus L_2$, then $s(a \circ L) = a \cdot s(L_1) + a \cdot s(L_2)$, similarly for $Q(a \circ L)$, $q(a \circ L)$ and $q_i(a \circ L)$. If $s(L) = \pi^i\mathfrak{o}$, then $L = L(i)$. In general $L \supseteq \cdots \supseteq L(i) \supseteq L(i+1) \supseteq \cdots \supseteq L(\infty) = K(L)$; $\pi^j L(i) \subseteq L(i+j)$; $kL(i) = kL$, when $i < \infty$; and $(L_1 \oplus L_2)(i) = L_1(i) \oplus L_2(i)$.

LEMMA 1.2. *Let L be a quadratic lattice. Then there exist $e \in k$ and $u \leqq v \leqq \infty$ such that $q(L) = e(\pi^u \mathfrak{o}^2 + \pi^v \mathfrak{o}^2)$. Moreover, they may be chosen to satisfy the following:*

1). $\operatorname{ord}(e) = 0$ *and* $e\pi^u \in Q(L)$,

2). *If* $q(L) = 0$, *then* $u = v = \infty$,

3). *If* $\operatorname{rank}_{\mathfrak{o}^2} q(L) = 1$, *then* $u < v = \infty$, *and* $s(L) = 0$,

4). *If* $\operatorname{rank}_{\mathfrak{o}^2} q(L) = 2$, *then* $u < v < \infty$, $u + v \equiv 1 \bmod 2$ *and* $v \leqq s + 1$, *where* $s(L) = \pi^s \mathfrak{o}$.

When these conditions hold, then u and v are uniquely determined by L. Such representations of $q(L)$ will be called standard.

Proof. We may assume that $q(L) \neq 0$. Thus, $s(L) = \pi^s \mathfrak{o} \subseteq q(L)\mathfrak{o} = \pi^u \mathfrak{o}$. If $u < s$ $(s < \infty)$, then there exists $x \in L$ such that $\operatorname{ord}(Q(x)) = u$. If $u = s$, then there exist $x, y \in L$ such that $\operatorname{ord}(\langle x, y\rangle) = u$ and one of the three elements $Q(x)$, $Q(y)$ and $Q(x+y)$ must have ordinal u. Thus, we can find $x \in L$ such that $Q(x) = e\pi^u$ with $\operatorname{ord}(e) = 0$. By scaling L, we may assume that $e = 1$. Thus, $\pi^u \mathfrak{o}^2 \subseteq q(L)$. In case 3), we see easily that $s(L) = 0$ and $q(L) = \pi^u \mathfrak{o}^2$. In case 4), it is clear that v may be found to satisfy all the conditions. Since $\mathfrak{o}^2$-modules of finite type in k have ranks at most 2, e, u and v may always be found to satisfy the conditions stated.

The uniqueness of u follows from the equation $\pi^u \mathfrak{o} = q(L)\mathfrak{o}$. Let $\operatorname{rank}_{\mathfrak{o}^2} q(L) = 2$, then v is the minimal ordinal in the set $\{\operatorname{ord}(a) \mid a \in q(L)$ and $\operatorname{ord}(a) \equiv u + 1 \bmod 2\}$. Thus, v is also uniquely determined, provided that it satisfies the conditions stated. Q.E.D.

A subset $\{x_i \mid 1 \leqq i \leqq m\}$ of a free $\mathfrak{o}$-module M of finite rank is called pure if it generates a direct summand of rank m. Clearly, given a set $\{y_i \mid 1 \leqq i \leqq m\}$ of k-independent elements in kM, there is a pure subset $\{x_i \mid 1 \leqq i \leqq m\}$ in M such that for $p = 1, 2, \cdots, m$, $\sum_{1 \leqq i \leqq p} kx_i = \sum_{1 \leqq i \leqq p} ky_i$.

Let M be a quadratic lattice and $-\infty < i \leqq \infty$. Then M is called an i-modular lattice if $\{x\} \subseteq M$ is a pure subset implies that $\langle x, M\rangle = \pi^i \mathfrak{o}$. If M is i-modular, then, $i = \infty$ implies that $M = K(M)$; $i < \infty$ implies that $K(M) = 0$, $s(M) = \pi^i \mathfrak{o}$ and $\{x\} \subseteq M$ is a pure subset if and only if $\langle x, M\rangle = \pi^i \mathfrak{o}$. If $M = M_1 \oplus M_2$, then M is i-modular if and only if M_1 and M_2 are i-modular.

LEMMA 1.3. *Let L be a quadratic lattice and J be a sublattice of L. Then,*

1). *If J is i-modular, then J is an orthogonal summand of L if and only*

if, $J \subseteq L(i)$ and J is a direct summand of $L(\infty)$ when $i = \infty$.

2). *If L is i-modular with $i < \infty$, then J is an orthogonal summand of L if and only if J is i-modular.*

Proof. 1) follows easily as in [15; Proposition 1, p. 160]. 2) follows from 1) and the preceding remarks.

COROLLARY 1.4. *Let M be an i-modular lattice, $i < \infty$, then*

$$M \cong \oplus_j \begin{pmatrix} & \pi^i \\ a_j & c_j \end{pmatrix}$$

and conversely.

Proof. By induction on $\dim M$ and Lemma 1.3.

2. Orthogonal decompositions and change of bases.

Definition 2.1. Let L be a quadratic lattice. By an easy induction argument, using Lemma 1.3, we obtain an orthogonal decomposition $L = \bigoplus_{1 \leqq i \leqq m} L_i \oplus \bigoplus_{1 \leqq i \leqq p} K_i$, where $\dim L_i = 2$, $\dim K_i = 1$, $0 \leqq p \leqq 2$, L_i is $s(i)$-modular with $s(1) \leqq \cdots \leqq s(m) < \infty$, and $K(L) = \bigoplus_{1 \leqq i \leqq p} K_i$. By Corollary 1.4, we have $L = \bigoplus_{1 \leqq i \leqq n} M_i \oplus K(L)$, where M_i is $t(i)$-modular with $t(1) < \cdots < t(n) < \infty$. These decompositions will be called complete and canonical respectively.

Let $L_i = \mathfrak{o}x_i + \mathfrak{o}y_i$ and $K_i = \mathfrak{o}z_i$ in the complete decomposition. Then, the following transformations will be called elementary:

LT_0: a). $z'_i = z_{i+1}$; $z'_{i+1} = z_i$,

b). $z'_i = z_i + az_{i+1}$, $\operatorname{ord}(a) \geqq 0$,

c). $z'_i = az_i$, $\operatorname{ord}(a) = 0$,

d). $x'_m = x_m + az_j$, $\operatorname{ord}(a) \geqq 0$,

LT_1: a). $x'_i = y_i$, $y'_i = x_i$,

b). $x'_i = x_i + ay_i$, $\operatorname{ord}(a) \geqq 0$,

c). $x'_i = ax_i$, $\operatorname{ord}(a) = 0$,

LT_2: a). $x'_i = x_{i+1}$, $y'_i = y_{i+1}$; $x'_{i+1} = x_i$, $y'_{i+1} = y_i$, provided that $s(i) = s(i+1) = \operatorname{ord}(\langle x_i, y_i \rangle) = \operatorname{ord}(\langle x_{i+1}, y_{i+1} \rangle)$,

b). $x'_i = x_i + ax_{i+1}$, $y'_i = y_i$; $x'_{i+1} = x_{i+1}$, $y'_{i+1} = ab_{i+1}b_i^{-1}y_i + y_{i+1}$, where $\operatorname{ord}(a) \geqq 0$, $b_i = \langle x_i, y_i \rangle$ and $b_{i+1} = \langle x_{i+1}, y_{i+1} \rangle$.

It is to be understood that in each of the nine types of transformations, all the remaining basis elements remain unchanged. Transformations obtained by interchanging the roles of x_i and y_i will be called elementary of the same type.

It is easy to see that the new basis elements in the given order will again give a complete decomposition of L.

LEMMA 2.2. *Let L be a quadratic lattice with a complete decomposition as in Definition 2.1. Let $\{x\} \subseteq L$ be a pure subset such that $\langle x, L\rangle = s(L)$. Then there is a sequence of elementary transformations which will lead from the given basis to a new basis such that x is the first basis element.*

Proof. If $m = 0$ or 1, then the hypothesis that $\langle x, L\rangle = s(L)$ together with transformations of type LT_0 and LT_1 easily give the desired result. Thus, we proceed by induction on m and let $m > 1$. By using transformations of type LT_1, LT_2 a) and the hypothesis that $\langle x, L\rangle = s(L)$, it is easy to see that we may assume that $x = x_1 + \sum_{2 \leqq i \leqq m} \alpha_i x_i + \sum_{1 \leqq i \leqq p} \gamma_i z_i$, where $\alpha_i, \gamma_i \in \mathfrak{o}$. Using transformations of type LT_2 b), we may first assume that $\alpha_2 = 0$, then $\alpha_2 = 1$. Thus, by induction, we may assume that $x = x_1 + x_2$ and another application of LT_2 b) gives the desired result. Q. E. D.

THEOREM 2.3. *Let x_i, y_i, z_j and x'_i, y'_i, z'_j be the bases elements of two complete decompositions of a quadratic lattice L. Then, there is a sequence of elementary lattice transformations which lead from the first to the second set of basis elements.*

Proof. If $m = 0$, then our assertion reduces to the classical result on changing basis in a free $\mathfrak{o}$-module of finite type. If $m > 0$, then by Lemma 2.2, we may let $y'_1 = y_1$. It is clear that $\mathrm{ord}(\langle x_1, y_1\rangle) = \mathrm{ord}(\langle x'_1, y'_1\rangle) < \infty$. Hence, we may assume that x'_1 has the form given in the proof of Lemma 2.2. We now observe that the transformations carried out in Lemma 2.2 do not alter y_1. Hence, we may assume that $x'_1 = x_1$. Now by induction, we get the desired result. Q. E. D.

3. Hyperbolic lattices.

Definition 3.1. A quadratic space U is called hyperbolic if it is the orthogonal sum of hyperbolic planes, $\begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix}_k$. If $i < \infty$, then a quadratic lattice L is called i-hyperbolic, when it is the orthogonal sum of i-hyperbolic planes, $\begin{pmatrix} & \pi^i \\ 0 & 0 \end{pmatrix}$. If M is any lattice, H is an i-hyperbolic plane, then $q_i(M \oplus H)$

$= q_i(M)$; in addition, if M is i-modular, then $Q(M \oplus H) = q(M \oplus H) = q(M) = q_i(M)$.

LEMMA 3.2. *A quadratic lattice L is i-hyperbolic if and only if, L is i-modular, kL is a hyperbolic space and $q(L) = \pi^i\mathfrak{o}$.*

Proof. The necessity is clear. Conversely, suppose that the three conditions hold. By Lemma 2.2, we may find $x, y \in L$ such that $M = \mathfrak{o}x + \mathfrak{o}y = \begin{pmatrix} & b \\ 0 & c \end{pmatrix}$, where $\operatorname{ord}(b) = i$, $\operatorname{ord}(c) \geqq i$. Replacing y by $\pi^i b^{-1}(cb^{-1}x + y) \in M$, we see that M is i-hyperbolic. By Lemma 1.3, $L = M \oplus M'$. By the cancellation theorem, [1; Satz 6, p. 157], kM' is hyperbolic. Thus, by Lemma 1.3 and Definition 3.1, we see that M' is i-modular and $q(M') = \pi^i\mathfrak{o}$, (provided that $M' \neq 0$.) Thus, by induction on $\dim L$, we get the desired result.

Q. E. D.

THEOREM 3.3. *Let $L = H_j \oplus L_j$, $j = 1, 2$, be two decompositions of a lattice L such that H_j is an i-hyperbolic plane for $j = 1, 2$. Then, $L_1 \cong L_2$.*

Proof. By scaling L, we may assume that $H_j = \mathfrak{o}x_j + \mathfrak{o}y_j = \begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix}$. Thus, $H_1 + H_2 \subseteq L(0)$ and $\langle H_1, H_2\rangle \subseteq \mathfrak{o}$.

Case 1. $\langle H_1, H_2\rangle = \mathfrak{o}$. Without loss of generality, we may assume that $\langle x_1, y_2\rangle = 1$. Thus, $H' = \mathfrak{o}x_1 + \mathfrak{o}y_2 = \begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix} \subseteq L(0)$. By Lemma 1.3, $L = H' \oplus L'$. Thus, by symmetry, we may assume that $x_1 = x_2$ and $y_1 \neq y_2$. Let z_j, $1 \leqq j \leqq m$, be a free $\mathfrak{o}$-basis for L_1. Then, $y_2 = ax_1 + by_1 + z$, where $z \in L_1$ and $1 = \langle x_2, y_2\rangle = \langle x_1, y_1\rangle = b$. Since $y_2 \in H_2 \subseteq L(0)$, we have $\langle y_2, z_j\rangle \in \mathfrak{o}$ and $w_j = \langle y_2, z_j\rangle x_1 + z_j \in L$ for $1 \leqq j \leqq m$. It is now trivial to see that $L = (\mathfrak{o}x_2 + \mathfrak{o}y_2) \oplus \sum_{1 \leqq j \leqq m} \mathfrak{o}w_j$. Thus, $L_2 = \sum_{1 \leqq j \leqq m} \mathfrak{o}w_j$. By computation, we see that the map $z_j \to w_j$ leads to an isometry between L_1 and L_2.

Case 2. $\langle H_1, H_2\rangle \subseteq \pi\mathfrak{o}$. Let $\{z_1 + z_2\}$ be a pure subset of $H = H_1 + H_2$, where $z_j \in H_j$. Without loss of generality, let $\{z_1\}$ be a pure subset of H_1. From $\langle H_1, H_2\rangle \subseteq \pi\mathfrak{o}$, $H \subseteq L(0)$, we see that $\mathfrak{o} = \langle z_1, H_1\rangle = \langle z_1 + z_2, H_1\rangle \subseteq \langle z_1 + z_2, H\rangle \subseteq \mathfrak{o}$. Thus, H is 0-modular. By Lemma 1.3, $L = H \oplus M$, $H = H_j \oplus M_j$; thus, we may assume that $L = H = H_1 + H_2 = H_j \oplus L_j$. Hence, $x_2 = w' + z'$ and $y_2 = w'' + z''$, where $w', w'' \in H_1$ and $z', z'' \in L_1$. Since $L = H_1 + H_2$, we see that $L_1 = \mathfrak{o}z' + \mathfrak{o}z''$. From $\langle H_1, H_2\rangle \subseteq \pi\mathfrak{o}$, we see that $w', w'' \in \pi H_1$. From $Q(x_2) = Q(y_2) = 0$ we get $Q(z') = Q(w')$ and $Q(z'') = Q(w'')$. Thus, $Q(z'), Q(z'') \in q(\pi H_1) = \pi^2\mathfrak{o}$. We have shown that

$H = L$ is 0-modular, thus, L_1 is 0-modular by Lemma 1.3, hence, $\operatorname{ord}(\langle z', z''\rangle) = 0$. By Lemma 1.1, $\Delta(kL_1) = 0$, thus, kL_1 is a hyperbolic space, cf. [1; Zusatz 2, p. 153]. It is also clear that $q(L_1) = \mathfrak{o}$. Hence, by Lemma 3.2, L is 0-hyperbolic. By symmetry, L_2 is also 0-hyperbolic. Hence, $L_1 \cong L_2$.

Q.E.D.

4. Isometry of modular lattices.

LEMMA 4.1. *Let* $B_j = \mathfrak{o}x_j + \mathfrak{o}y_j = \begin{pmatrix} \pi^i & \\ a & c_j \end{pmatrix}$ *and* $q(B_j) \subseteq a\mathfrak{o}$, $j = 1, 2$. *Then* $B_1 \cong B_2$ *if and only if* $\Delta(kB_1) = \Delta(kB_2)$.

Proof. The necessity is obvious. Conversely, by scaling, we may assume that $i = 0$. Since $\mathfrak{o} \subseteq q(B_j) \subseteq a\mathfrak{o}$, we have $\operatorname{ord}(a) \leqq 0$ and $\operatorname{ord}(c_1 + c_2) \geqq \operatorname{ord}(a)$. Let $d \in k$ such that $a(c_1 + c_2) = \mathfrak{P}(da)$. Then $c_2 = c_1 + d^2a + d$ and $\operatorname{ord}(d) + \operatorname{ord}(1 + da) = \operatorname{ord}(c_1 + c_2) \geqq \operatorname{ord}(a)$. If $\operatorname{ord}(d) < 0$, then $\operatorname{ord}(d) + \operatorname{ord}(1 + da) < \operatorname{ord}(1 + da) = \operatorname{ord}(da) < \operatorname{ord}(a)$, a contradiction. Thus, $\operatorname{ord}(d) \geqq 0$, $dx_1 + y_1 \in B_1$ and we have $Q(dx_1 + y_1) = d^2a + d + c_1 = c_2$. Hence $B_1 \cong B_2$.

Q.E.D.

COROLLARY 4.2. *Let* $B \cong \begin{pmatrix} 1 & \\ \pi^u & c \end{pmatrix}$ *be a quadratic lattice such that* $q(B) \subseteq \pi^u\mathfrak{o}$. *Let* $\pi^u c = \pi^{-1}b^2 + D + \mathfrak{P}(d)$ *as in Lemma* 1.1, 5). *Then* $B \cong \begin{pmatrix} 1 & \\ \pi^u & \pi^{-u}(\pi^{-1}b^2 + D) \end{pmatrix}$.

Proof. By Lemma 4.1 and Lemma 1.1, 5).

LEMMA 4.3. *Let* M *be an* i*-modular lattice,* $i < \infty$. *Then,* $\pi^{-i}\circ M \cong \begin{pmatrix} 1 & \\ 1 & D \end{pmatrix} \oplus H$, *where* $D = 0$ *or* λ *and* H *is* 0 *or* 0*-hyperbolic, if and only if,* $q(M) = \pi^i\mathfrak{o}$.

Proof. The necessity is obvious. By scaling, we may take $i = 0$. From Corollary 1.4 and that $q(M) = \mathfrak{o}$, we have $M \cong \bigoplus_{1 \leqq j \leqq m} \begin{pmatrix} 1 & \\ a_j & c_j \end{pmatrix}$, where $a_j, c_j \in \mathfrak{o}$, $1 \leqq j \leqq m$. If $a_jc_j \in \pi\mathfrak{o}$, then by Lemma 1.1 and Lemma 3.2, $\begin{pmatrix} 1 & \\ a_j & c_j \end{pmatrix} \cong \begin{pmatrix} 1 & \\ 0 & 0 \end{pmatrix}$. If $\operatorname{ord}(a_j) = \operatorname{ord}(c_j) = 0$, then an easy application of Lemma 4.1 and Lemma 1.1 shows that we may assume $c_j \neq 0 \in \Omega$. By symmetry, we may assume that $a_j \neq 0 \in \Omega$. Using a transformation of type LT_1 c), we may assume that $a_j = 1$. By Corollary 4.2, we may replace c_j by 0 or λ. As we have just seen $c_j = 0$ leads to a 0-hyperbolic plane. By using a transformation of type LT_2 b), $\begin{pmatrix} 1 & \\ 1 & \lambda \end{pmatrix} \oplus \begin{pmatrix} 1 & \\ \lambda & 1 \end{pmatrix} \cong \begin{pmatrix} 1 & \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 1 & \\ 0 & 1 \end{pmatrix}$. From this we get our assertion immediately.

Q.E.D.

LEMMA 4.4. *Let M be an i-modular lattice, $i < \infty$. If* $\dim M \geqq 6$, *then M contains an i-hyperbolic plane.*

Proof. We may assume that $\dim M = 6$ and $i = 0$. Let $q(M) = e(\pi^u \mathfrak{o}^2 + \pi^v \mathfrak{o}^2)$ be a standard representation of the norm group. Since $\operatorname{ord}(e) = 0$, by scaling, we may assume that $e = 1$ and $\pi^u = Q(x)$ for some $x \in M$, $u < v \leqq 1$, and $u + v \equiv 1 \bmod 2$. $\{x\}$ is a pure subset of M; hence, we can find $y \in M$ such that $\operatorname{ord}(\langle x, y\rangle) = 0$. From Lemma 1.3, $M = (\mathfrak{o}x + \mathfrak{o}y) \oplus N$. By the choice of x and using transformations of type LT_0 d) on $\mathfrak{o}x \oplus N$, we may assume that $\mathfrak{o}x \oplus N = (\pi^u) \oplus \begin{pmatrix} & 1 \\ a & b \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ c & d \end{pmatrix}$, where $a, b, c, d \in \pi^v \mathfrak{o}^2$, and $\operatorname{ord}(a) \leqq \operatorname{ord}(c) \leqq \operatorname{ord}(d)$. By transformations of type LT_0 d) on $(a) \oplus \begin{pmatrix} & 1 \\ c & d \end{pmatrix}$, we see that M contains a 0-hyperbolic plane.

Q. E. D.

LEMMA 4.5. *Let M be an i-module lattice, $i < \infty$, and* $\dim M \geqq 4$. *Let $q(M) = e(\pi^u \mathfrak{o}^2 + \pi^v \mathfrak{o}^2)$ be standard and $e\pi^u = Q(x)$. Then,*

1). $q(M) = Q(M)$.

2). $e^{-1}\pi^{-i} \circ M \cong \begin{pmatrix} & 1 \\ a_1 & c_1 \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ a_2 & c_2 \end{pmatrix} \oplus H$, *where H is 0-hyperbolic or zero and the following hold:*

Case 1. $v \leqq i$. $a_1 = \pi^{u-i}$, $c_1 = \pi^{-u+i} D_1 + \pi^{v+i} b^2$, $a_2 = \pi^{v-i}$, *and* $c_2 = \pi^{-v+i} D_2$, *where* $D_1, D_2 \in \{0, \lambda\}$ *and* $\operatorname{ord}(b) \geqq 0$.

Case 2. $v = i + 1$ *and* $u < i$. a_1 *and* c_1 *as in Case* 1, $a_2 = c_2 = 0$.

Case 3. $v = i + 1$ *and* $u = i$. $a_1 = 1$, $c_1 = D_1$, $a_2 = c_2 = 0$.

Proof. 1) follows from 2).

2). By scaling, we may assume that $i = 0$ and $e = 1$. By Lemma 4.4 and Definition 3.1, we may assume that $\dim M = 4$. $\{x\}$ is clearly a pure subset of M, thus, by Lemma 1.3 and Lemma 2.2, we have $M \cong \begin{pmatrix} & 1 \\ \pi^u & f \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ g & h \end{pmatrix}$. Using transformations of type LT_2 b), we may assume that $g, h \in \pi^v \mathfrak{o}^2$.

Case 3 now follows easily from Lemma 4.3.

Case 2. $v = 1$ and $u < 0$. By Lemma 3.2 (also the proof of Lemma 4.3), we may assume that $g = h = 0$. By Corollary 4.2, we may assume that $f = \pi^{-u}(\pi^{-1} b_1^2 + D)$, where $b_1 \in \Omega[\pi^{-1}]$ and $D = 0$ or λ. If $b_1 \neq 0$, then $\operatorname{ord}(f) + u \equiv 1 \bmod 2$, hence f has the desired form as c_1.

Case 1. Let $v=0$. By Lemma 4.3, we may assume that $g=1$ and $h=0$ or λ. Repeating Case 2, we get the desired result.

Let $v<0$ and $\operatorname{ord}(g)=v$. Using transformations of type LT_1 c), we may assume that $g=\pi^v$. As in Case 2, we may replace h by $\pi^{-v}(\pi^{-1}b_1^2+D)$. By choice, $\pi^{-v-1}b_1^2 \in \pi^u \mathfrak{o}^2$. Thus, using a transformation of type LT_2 b), we may assume that $b_1=0$. One more application of Corollary 4.2 on $\begin{pmatrix} & 1 \\ \pi^u & f \end{pmatrix}$ gives the desired form.

Finally, let $v<0$ and $v<\operatorname{ord}(g)\leqq\operatorname{ord}(h)$. Let $M'=\mathfrak{o}x+\mathfrak{o}y$ $=\begin{pmatrix} & 1 \\ \pi^u & f \end{pmatrix}\subseteq M$. Since $v<0$, we have $q(M')=q(M)$. Thus, for some $\alpha,\beta\in\mathfrak{o}$, $\operatorname{ord}(Q(\alpha x+\beta y))=v$. Clearly, $\{\alpha x+\beta y\}$ is a pure subset of M'. If $\operatorname{ord}(\beta)>0$, then $\operatorname{ord}(\alpha)=0$; hence,

$$\operatorname{ord}(Q(\alpha x+\beta y))=\operatorname{ord}(\alpha^2\pi^u+\alpha\beta+\beta^2 f)=u<v<0,$$

a contradiction. Thus, $\operatorname{ord}(\beta)=0$. Replacing y by $\beta^{-1}(\alpha x+\beta y)$, we may assume that $\operatorname{ord}(f)=v$. Using a transformation of type LT_2 b), we have $M\cong\begin{pmatrix} & 1 \\ \pi^u+g & f \end{pmatrix}\oplus\begin{pmatrix} & 1 \\ g & h+f \end{pmatrix}$, where $h+f=\pi^v f_1^2$, $\operatorname{ord}(f_1)=0$. Using transformations of type LT_1, $M\cong\begin{pmatrix} & 1 \\ \pi^u+g & f \end{pmatrix}\oplus\begin{pmatrix} & 1 \\ \pi^v & h' \end{pmatrix}$, where $g,h'\in\pi^v\mathfrak{o}^2$. Using a transformation of type LT_2 b), we have $M\cong\begin{pmatrix} & 1 \\ \pi^u & f' \end{pmatrix}\oplus\begin{pmatrix} & 1 \\ \pi^v & h'' \end{pmatrix}$. Using one more transformation of type LT_2 b), we may assume that $h''\in\pi^v\mathfrak{o}^2$. Repeating the argument of the preceding paragraph, we get the desired result.

Q.E.D.

THEOREM 4.6. *Let M_1 and M_2 be i-modular lattices. Then, $M_1\cong M_2$ if and only if $kM_1\cong kM_2$ and $q(M_1)=q(M_2)$.*

Proof. The necessity is obvious. If $i=\infty$, then $q(M_j)=Q(M_j)$, $j=1,2$. It is easy to see that either $M_j\cong(e\pi^u)$ or $M_j\cong(e\pi^u)\oplus(e\pi^v)$, where $q(M_j)$ $=e(\pi^u\mathfrak{o}^2+\pi^v\mathfrak{o}^2)$ is a standard representation.

Thus, let $i<\infty$, $kM_1\cong kM_2$, $q(M_1)=q(M_2)$ and M_1, M_2 be i-modular. By Definition 3.1 and the cancellation theorem, cf. [1; Zusatz 2, p. 153], these conditions are "invariant" under addition or deletion of i-hyperbolic planes. Thus, by Theorem 3.3 and Lemma 4.4, we may assume that $\dim M_1=\dim M_2$ $=4$. By scaling, we may assume that $i=0$ and $q(M_j)=\pi^u\mathfrak{o}^2+\pi^v\mathfrak{o}^2$, $u<v\leqq 1$, $u+v\equiv 1 \bmod 2$, $j=1,2$. If $v=1$, then by Lemma 4.5, 2), Case 3 and a simple computation of the Arf invariant of kM_1 and kM_2, we see that $M_1\cong M_2$. Thus, we may assume that $v\leqq 0$. By Corollary 4.2 and Lemma 4.5, 2),

$$M_j \cong \begin{pmatrix} & 1 \\ \pi^u & \pi^{-u}(\pi^{-1}b_j^2 + D'_j) \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ \pi^v & \pi^{-v}D_j \end{pmatrix},$$

where $b_j \in \Omega[\pi^{-1}]$, $D'_j, D_j \in \{0, \lambda\}$, $j = 1, 2$.

If u is even, then v is odd. Thus, $[C(kM_j)] = [\pi^v, \pi^{-v}D_j] = [\pi^{-1}, \pi D_j]$ for $j = 1, 2$. Hence, from $kM_1 \cong kM_2$, we get $D_1 = D_2$. From Lemma 4.1 and that $\Delta(kM_1) = \Delta(kM_2)$, we see that $M_1 \cong M_2$.

If u is odd, then v is even. Thus, $[C(kM_j)] = [\pi^u, \pi^{-u}D'_j] = [\pi^{-1}, \pi D'_j]$ for $j = 1, 2$. As above, $D'_1 = D'_2$. From $kM_1 \cong kM_2$, we get $0 = \Delta(kM_1) + \Delta(kM_2) = \pi^{-1}(b_1 + b_2)^2 + (D_1 + D_2) + \mathfrak{P}k$. By Lemma 1.1, $b_1 = b_2$ and $D_1 = D_2$. Thus, $M_1 \cong M_2$. Q.E.D.

5. Isometry of lattices.

Definition 5.1. Let $L = \bigoplus_{0 \leq i \leq m} L_i \oplus K(L)$ be a canonical decomposition, where L_j is $s(j)$-modular, $s(0) < \cdots < s(m) < \infty$. It is easy to see that for i such that $s(n) \leq i < s(n+1)$, we have

$$L(i) = \bigoplus_{0 \leq j \leq n} \pi^{i-s(j)} L_j \oplus \bigoplus_{n+1 \leq j \leq m} L_j \oplus K(L).$$

Let $q_i(L) = e_i(\pi^{u(i)}\mathfrak{o}^2 + \pi^{v(i)}\mathfrak{o}^2)$ be standard representations of the i-th norm group for $i \leq \infty$, i.e. they are standard representations of $q(L(i) \oplus H_i)$ or of $q(K(L))$, where H_i is an i-hyperbolic plane, when $i < \infty$, we do not assert that $e_i\pi^{u(i)} \in Q(L(i))$. Clearly, we have $q_i(L) \supseteq q_{i+1}(L) \supseteq \cdots \supseteq q_\infty(L)$ and $\pi^2 q_i(L) \supseteq q_{i+1}(L)$. From these we deduce easily the following possibilities:

1). $u(j+1) = u(j)$. Then, $v(j+1) = v(j)$ or $v(j) + 2$ and $= v(j)$ when $v(j) = j+1$, and $q_j(L) = e_{j+1}(\pi^{u(j)}\mathfrak{o}^2 + \pi^{v(j)}\mathfrak{o}^2)$.

2). $u(j+1) = u(j) + 1$. Then, $v(j+1) = v(j) + 1 = u(j+1) + 1$ and $q_j(L) = \pi^{u(j)}\mathfrak{o}$, $q_{j+1}(L) = \pi^{u(j)+1}\mathfrak{o} = \pi^{u(j+1)}\mathfrak{o}$.

3). $u(j+1) = u(j) + 2$. Then, $q_{j+1}(L) = e_j(\pi^{u(j+1)}\mathfrak{o}^2 + \pi^{v(j)}\mathfrak{o}^2)$.

The given canonical decomposition is said to be saturated in the $s(j)$-th component, if $q_{s(j)}(L) = Q(L_j) = q(L(s(j)))$. It is said to be saturated, if it is saturated in each $s(i)$-th component, $0 \leq i \leq m$.

If $L = K(L)$, then let $F(L) = T(L) = \infty$ and $D(L) = (\infty)$. If $L \neq K(L)$, then let $F(L) = s(0)$, $T(L) = s(m) + n$, where n is the largest finite integer among 0, $u(\infty) - u(s(m))$ and $u(\infty) - u(s(m)) + v(\infty) - v(s(m))$, and $D(L) = (r_{F(L)}, \cdots, r_{T(L)})$, where $r_i = 0$ unless $i = s(j)$, then $r_i = \dim L_j$. $(F(L), T(L))$ and $D(L)$ are called the type and the type dimension of L respectively. L is called a normal lattice when $r_i \geq 8$ for

$F(L) \leqq i \leqq T(L)$. Thus, either $L = K(L)$ or when $L \neq K(L)$, then $u(\infty) = u(T)$, $v(\infty) = v(T)$ or ∞, and $s(j) = F + j$, $0 \leqq j \leqq T - F$. Clearly, $F(L)$, $T(L)$ and $D(L)$ are invariants of L.

LEMMA 5.2. *Let $L = \bigoplus_{0 \leqq i \leqq m} L_i \oplus K(L)$ be a canonical decomposition of the lattice L of type (F,T). Let $H = 0$ when $F = T = \infty$ and $H = \bigoplus_{F \leqq j \leqq T} H_j$, where H_j is j-hyperbolic of dimension 8 otherwise. Then,*

1). *$q_i(L \oplus H) = q_i(L)$ for $i = F, F+1, \cdots, T, \infty$.*

2). *$L \oplus H$ is a normal lattice of type (F,T).*

Proof. If $L = K(L)$, then there is nothing to prove. Thus, we assume that $(F,T) \neq (\infty, \infty)$.

1). follows from the definition of $q_i(L)$.

2). If $K(L) = 0$, then the normality follows easily from 1). Thus, let $K(L) \neq 0$ and n be chosen such that $T(L) = s(m) + n = T$.

From Definition 5.1 and 1), $q_T(L \oplus H) = q(\pi^n L(s(m))) + q_\infty(L) + \pi^T \mathfrak{o}$. Since $L_m \neq 0$ is $s(m)$-modular, let

$$q(L(s(m))) = q_{s(m)}(L) = e_{s(m)}(\pi^{u(s(m))}\mathfrak{o}^2 + \pi^{v(s(m))}\mathfrak{o}^2)$$

be a standard representation. By the choice of n, Definition 5.1 and that $K(L) \neq 0$, we see that $T = s(m) + n \geqq u(\infty) + [s(m) - u(s(m))] \geqq u(\infty)$ and $2n + u(s(m)) \geqq u(\infty) + [u(\infty) - u(s(m))] \geqq u(\infty)$. Thus, $q_T(L \oplus H)\mathfrak{o} = q_\infty(L)\mathfrak{o}$, or $u(T) = u(\infty)$. If $v(\infty) = \infty$, then by 1) and definition, $L \oplus H$ is normal. Let $v(\infty) < \infty$. If $u(\infty) \geqq v(s(m))$, then $n \geqq v(\infty) - u(s(m))$, thus, we get $T \geqq v(\infty)$ and $2n + u(s(m)) \geqq v(\infty)$. From $\pi^{v(\infty)}\mathfrak{o} \subseteq q_\infty(L)$ we get $q_T(L \oplus H) = q_\infty(L)$, thus $L \oplus H$ is normal by 1). Finally, if $u(\infty) < v(s(m))$, then $u(s(m)) \equiv u(\infty)$ and $v(s(m)) \equiv v(\infty)$ mod 2. The argument used in case $v(\infty) = \infty$ shows that we may assume $u(s(m)) = u(\infty)$. Hence, we may assume that $e_{s(m)} = e_\infty$. It is now trivial to see that $q_T(L \oplus H) = q_\infty(L)$, thus, $L \oplus H$ is normal. Q. E. D.

LEMMA 5.3. *A normal lattice L admits a saturated canonical decomposition.*

Proof. Let $L = \bigoplus_{0 \leqq i \leqq m} L_i \oplus K(L)$ be a canonical decomposition. If $L = K(L)$, then the decomposition is already saturated. Thus, we may assume that $L \neq K(L)$. By scaling, we may assume that L has type $(0, m)$. If $m = 0$ and $K(L) = 0$, then Lemma 4.5 furnishes the desired result. If

$m=0$ and $K(L)\neq 0$, then by Lemma 4.4 and Lemma 4.5, we may assume that $L=L'_0\oplus H\oplus K(L)$, where H is 0-hyperbolic of dimension 4 and $Q(L'_0)=q(L_0)$. Let $q(K(L))=e(\pi^u 0^2+\pi^v 0^2)$ be a standard representation. Then by using transformations of type LT_0 on $H\oplus K(L)$, we have $H\oplus K(L)\cong H'\oplus K(L)$, where $H'\cong\begin{pmatrix} & 1\\ e\pi^u & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ e\pi^v & 0\end{pmatrix}$, since $Q(K(L))=q(K(L))$. Thus, $L=L'\oplus K(L)$, where $L'\cong L'_0\oplus H'$. It is clear that $q(L')=q(L)=q_0(L)$ and $L\cong L'\oplus K(L)$ is a saturated canonical decomposition. Hence, we may assume that $m>0$.

By Lemma 4.4 and Lemma 4.5, let $L_i=L'_i\oplus H_i$, where H_i is i-hyperbolic of dimension 4 and $q(L_i)=Q(L'_i)$, $i=0,1$. By Theorem 4.6, $L_0\cong L'_0\oplus H'_0$, where $H'_0\cong\begin{pmatrix} & 1\\ a & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ c & 0\end{pmatrix}$, $a,c\in q(L_0)$. By using transformations of type LT_2, we see that $H'_0\oplus H_1\cong H'_0\oplus H'_1$, where

$$H'_1\cong\begin{pmatrix} & 1\\ \pi^2 a & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ \pi^2 c & 0\end{pmatrix}.$$

Thus, from Definition 5.1, we see that we may assume that $L=L_0\oplus L'$ with $q_i(L')=q_i(L)$, $i=1,\cdots,m,\infty$. Hence, L' is a normal lattice of type $(1,m)$. By induction, we may then assume that the chosen canonical decomposition is already saturated in the i-th component, $1\leqq i\leqq m$. Now, by the same argument, $L_1\cong L'_1\oplus H''_1$, where $H''_1\cong\begin{pmatrix} & \pi\\ b & 0\end{pmatrix}\oplus\begin{pmatrix} & \pi\\ d & 0\end{pmatrix}$ and $H_0\oplus H''_1\cong H''_0\oplus H''_1$, where $H''_0\cong\begin{pmatrix} & 1\\ b & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ d & 0\end{pmatrix}$ and $b,d\in q_1(L)$ are arbitrary. Thus, we can saturate the 0-th component without destroying the saturation of the remaining components. Hence our assertion holds.

Q. E. D.

LEMMA 5.4. *Let* $L=\bigoplus_{0\leqq i\leqq m}L_i\oplus K(L)=\bigoplus_{0\leqq i\leqq m}M_i\oplus K(L)$ *be two canonical decompositions of the normal lattice* L *of type* (F,T). *Let* $q_i(L)=e_i(\pi^{u(i)}0^2+\pi^{v(i)}0^2)$, $i=F,F+1,\cdots,T,\infty$, *be standard representations, cf. Definition* 5.1. *Then,*

1). $\operatorname{Ord}\left(\sum_{0\leqq j\leqq i-1}(\Delta(kL_j)+\Delta(kM_j))\right)$
$$\geqq u(F+i)+v(F+i)-(2F+2i+1),$$

2). *If* $v(F+i)=F+i+1$, *then*
$$\bigoplus_{0\leqq j\leqq i-1}kL_j\oplus(e_{F+i}\pi^{u(F+i)})_k\cong\bigoplus_{0\leqq j\leqq i-1}kM_j\oplus(e_{F+i}\pi^{u(F+i)})_k,$$

where $1\leqq i\leqq T-F=m$. ($\infty-\infty$ *is taken to mean* 0.)

Proof. Clearly, the assertions are "invariant" under scalings of L. By Theorem 2.3, we may assume that the given decompositions are related by an elementary lattice transformation. By inspection, it is clear that we need only consider transformations of type LT_2b). By scaling L and using transformations of type LT_1, we may assume that the given decompositions of L differ as follows:

$$\begin{pmatrix} 1 & \\ a & b\end{pmatrix} \oplus \begin{pmatrix} \pi & \\ c & d\end{pmatrix} \to \begin{pmatrix} 1 & \\ a+cg^2 & b\end{pmatrix} \oplus \begin{pmatrix} \pi & \\ c & g^2\pi^2 b + d\end{pmatrix}, \ \operatorname{ord}(g) \geqq 0;$$

it is to be understood that the given canonical decompositions are, in reality, complete decompositions, and the lattices above are orthogonal summands of the respective complete decompositions.

1). It suffices to show that $\operatorname{Ord}(bcg^2 + \mathfrak{P}k) \geqq u(1) + v(1) - 3$. From Definition 5.1, we see that $bcg^2 \in q_0(L) q_1(L) = \pi^{u(0)+u(1)}\mathfrak{o}^2 + \pi^{u(0)+v(1)}\mathfrak{o}^2 + \pi^{v(0)+u(1)}\mathfrak{o}^2 + \pi^{v(0)+v(1)}\mathfrak{o}^2$. If $u(1) = u(0)+1$, then by 2) of Definition 5.1, $\operatorname{ord}(bcg^2) \geqq u(0) + u(1) \geqq u(1) + v(1) - 3$. Thus, we may assume that $u(1) - u(0), v(1) - v(0) \in \{0, 2\}$. Hence, $bcg^2 = \pi^{u(0)+u(1)} f^2 + h$, where $\operatorname{ord}(f) \geqq 0$ and $\operatorname{ord}(h) \geqq u(1) + v(1) - 3$. Let $n = \frac{1}{2}(u(0) + u(1)) \geqq u(1) - 1$, $b' = bcg^2 + \mathfrak{P}(\pi^n f)$. Since $v(1) \leqq 2$, we have $n \geqq u(1) + v(1) - 3$ and $\operatorname{Ord}(bcg^2 + \mathfrak{P}k) \geqq \operatorname{ord}(b') = \operatorname{ord}(\pi^n f + h) \geqq u(1) + v(1) - 3$.

2). By scaling, we may assume that $e_1 = 1$. By the cancellation theorem [1; Folgerung, p. 160], it suffices to show that,

$$\begin{pmatrix} 1 & \\ a & b\end{pmatrix}_k \oplus (\pi^{u(1)})_k \cong \begin{pmatrix} 1 & \\ a+cg^2 & b\end{pmatrix}_k \oplus (\pi^{u(1)})_k$$

when $v(1) = 2$. From [1; Satz 8, p. 161], it is enough to show that $[\pi^{-u(1)} cg^2, \pi^{u(1)} b] = 1$. From Definition 5.1 and that $v(1) = 2$, we see that $v(0) = 0$ or 1 (note that $v(0) \leqq 1$ by definition).

If $v(0) = 1$, then $u(1) = 1$ and $b \in \mathfrak{o}$, $cg^2 \in \pi\mathfrak{o}$. Thus, $[\pi^{-1} cg^2, \pi b] = [C(V)]$, where $V = \begin{pmatrix} 1 & \\ \pi^{-1}cg^2 & \pi b\end{pmatrix}_k$. By Lemma 1.1, $\Delta(V) = 0$, thus V is a hyperbolic plane [1; Zusatz 2, p. 153], hence $[C(V)] = 1$ as observed in Section 1.

If $v(0) = 0$, then $u(1) = u(0)$ or $u(0) + 2$. If $u(1) = u(0)$, then we may take $e_0 = 1$, by using Definition 5.1, 1). If $u(1) = u(0) + 2$, then by Definition 5.1, we have $\pi^2 q_0(L) = e_0 q_1(L) \subseteq q_1(L)$; now, from $\operatorname{ord}(e_0) = 0$, we have $q_1(L) = e_0^2 q_1(L) \subseteq e_0 q_1(L)$; hence, $q_0(L) = \pi^{-2} q_1(L)$ and we may take $e_0 = 1$. Thus, in all cases, $b \in \pi^{u(0)}\mathfrak{o}^2 + \pi^{v(0)}\mathfrak{o}^2$ and $cg^2 \in \pi^{u(1)}\mathfrak{o}^2 + \pi^{v(1)}\mathfrak{o}^2$, where $v(1) = 2$ and $v(0) = 0$. By choice, $u(0)$ and $u(1)$ must both be odd.

Thus, $[\pi^{-u(1)}cg^2, \pi^{u(1)}b] = [\pi^{-u(1)+v(1)}c_2^2, \pi^{u(1)+v(0)}b_2^2]$, where $b_2, c_2 \in \mathfrak{o}$. Using the same argument of the preceding paragraph, we get the desired result.

Q. E. D.

THEOREM 5.5. *Let* $L = \bigoplus_{0 \leqq i \leqq m} L_i \oplus K(L)$ *and* $M = \bigoplus_{0 \leqq i \leqq n} M_i \oplus K(M)$ *be canonical decompositions. Then* $L \cong M$ *if and only if,*

1). $kL \cong kM$.

2). *L and M have the same type* (F, T) *and* $D(L) = D(M)$. *In particular,* $m = n \leqq T - F$ *when* $(F, T) \neq (\infty, \infty)$.

3). $q_i(L) = q_i(M) = e_i(\pi^{u(i)}\mathfrak{o}^2 + \pi^{v(i)}\mathfrak{o}^2)$, $i = F, \cdots, T, \infty$, *where the representations are standard, cf. Definition* 5.1.

4). $$\operatorname{Ord}\Big(\sum_{0 \leqq j \leqq p(i)} (\Delta(kL_j) + \Delta(kM_j))\Big) \geqq u(F+i) + v(F+i) - (2F + 2i + 1),$$

where $p(i)$ *satisfies* $s(p(i)) < F + i \leqq s(p(i)+1)$, L_j *is* $s(j)$*-modular and* $1 \leqq i \leqq T - F$.

5). *If* $v(F+i) = F + i + 1$, *then*

$$\bigoplus_{0 \leqq j \leqq p(i)} kL_j \oplus (e_{F+i}\pi^{u(F+i)})_k \cong \bigoplus_{0 \leqq j \leqq p(i)} kM_j \oplus (e_{F+i}\pi^{u(F+i)})_k,$$

where $p(i)$ *is as in* 4), *and* $1 \leqq i \leqq T - F$.

Proof. Using Theorem 3.3, we see that our assertion is invariant under simultaneous addition of i-hyperbolic planes to L and M, provided that $F \leqq i \leqq T$. Thus, by Lemma 5.2, we may assume that L and M are both normal lattices.

Necessity. By Lemma 5.4.

Sufficiency. A_1). By the necessity above and Lemma 5.3, we may assume that the given decompositions are saturated.

A_2). By 1) and 3), we may assume that

$$K(L) \cong K(M) \cong (e_\infty \pi^{u(\infty)}) \oplus (e_\infty \pi^{v(\infty)}),$$

where (0) is understood to be the zero lattice. In particular, we may assume that $(F, T) \neq (\infty, \infty)$.

A_3). By Lemma 4.4, we may assume that $L_i = L'_i \oplus H_i$, where $\dim L'_i = 4$ and H_i is i-hyperbolic. Using Lemma 4.5, we may assume that L'_i is

presented in the form stated in 2) of Lemma 4.5. By A_1), the quantities e, u and v of Lemma 4.5 correspond to e_i, $u(i)$ and $v(i)$ respectively. From Theorem 4.6, we see that $L_i \cong L'_i \oplus H'_i \oplus H''_i$, where

$$H'_i \cong \begin{pmatrix} & \pi^i \\ e_i\pi^{u(i)} & 0 \end{pmatrix} \left(\oplus \begin{pmatrix} & \pi^i \\ e_i\pi^{v(i)} & 0 \end{pmatrix} \right)$$

and H''_i is either zero or i-hyperbolic. Similarly for M_i.

A_4). We may always consider $a \circ L$ and $a \circ M$ in place of L and M, since the given conditions imply a similar set of conditions for the scaled lattices and the latter will also be normal and satisfy A_1), A_2) and A_3). In particular, we may assume that $(F,T) = (0,m)$ and proceed by induction on $T - F = m$.

A_5). $(F,T) = (0,0)$. Then, the following cases occur:

Case 1. $K(L) = K(M) = 0$. By Theorem 4.6, $L \cong M$.

Case 2. $K(L) \cong K(M) \cong (e_\infty \pi^{u(\infty)})$, $u(\infty) < \infty$.

By normality, $u(\infty) = u(0)$. By scaling, we may assume that $e_\infty = e_0 = 1$. From $u(0) \leqq 0$, Lemma 4.5 and using transformations of type LT_0 d) on $L'_0 \oplus K(L)$, we have,

$$L'_0 \oplus K(L) \cong \begin{pmatrix} & 1 \\ 0 & c \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ \pi^{v(0)} & \pi^{-v(0)}D \end{pmatrix} \oplus K(L), \text{ if } v(0) \leqq 0,$$

$$L'_0 \oplus K(L) \cong \begin{pmatrix} & 1 \\ 0 & c \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix} \oplus K(L), \text{ if } u(0) < 0 \text{ and } v(0) = 1,$$

$$L'_0 \oplus K(L) \cong \begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ 0 & 0 \end{pmatrix} \oplus K(L), \text{ if } u(0) = 0 \text{ and } v(0) = 1,$$

where $c \in \pi^{v(0)}\mathfrak{o}^2$ and $D = 0$ or λ. Applying Theorem 4.6 to $\begin{pmatrix} & 1 \\ 0 & c \end{pmatrix} \oplus H_0'$, we may assume that $c = 0$. Since $u(0) + v(0) \equiv 1 \bmod 2$ and $u(\infty) = u(0)$, we see that $[\pi^{-1}, \pi D] = [C(\pi^{u(\infty)} \circ kL)/\mathfrak{R}]$, where $\mathfrak{R}$ is the radical of the Clifford algebra $C(\pi^{u(\infty)} \circ kL)$. By condition 1) and interchanging L and M, we see that $L \cong M$.

Case 3. $K(L) \cong K(M) \cong (e_\infty \pi^{u(\infty)}) \oplus (e_\infty \pi^{v(\infty)})$, $v(\infty) < \infty$. By normality, $u(\infty) = u(0)$ and $v(\infty) = v(0)$. Thus, we see easily that $q_0(L) = q_\infty(L) = q_\infty(M) = q_0(M)$. By using transformations of type LT_0 d), we may assume that L'_0 and M'_0 are 0-hyperbolic. Thus, $L \cong M$.

A_6). $(F,T) = (0,m)$, $m > 0$. We assert that $L'_0 \oplus H'_1$ and $M'_0 \oplus H'_1$ contain isometric 0-modular sublattices of dimension 4, say, $L''_0 \cong M''_0$. We have the following cases to consider:

Case 1. $v(0)=v(1)$. Since $v(0)\leqq 1$, condition 5) is vacuous on the 0-modular components. Thus, by Definition 5.1 and scaling by e_0 or e_1, we may assume that $e_0=e_1=1$. Now $L'_0\oplus H'_1$ contains the sublattice $N=L'_0\oplus(\pi^{u(1)})\oplus(\pi^{v(1)})$. By Lemma 4.5 and using transformations of type LT_0 d) on N, we see easily that N contains $\begin{pmatrix} & 1\\ \pi^{u(0)} & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}$. Interchanging L and M, we get A_6).

Case 2. $v(0)+1=v(1)$. By Definition 5.1, $v(0)=u(0)+1=u(1)$ and scaling by e_1 allows us to assume that $e_0=e_1=1$. If $v(0)\leqq 0$, then the argument of Case 1 may be used to get A_6). Thus, we may assume that $v(0)=1$, hence $v(1)=2$. By Lemma 4.5, we see that $L'_0\cong\begin{pmatrix} & 1\\ 1 & D\end{pmatrix}\oplus\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}$ and $M'_0\cong\begin{pmatrix} & 1\\ 1 & D'\end{pmatrix}\oplus\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}$, where $D,D'\in\{0,\lambda\}$. Condition 5), [1; Satz 8, p. 161] and the remarks preceding Lemma 1.1 show that $D=D'$. Thus A_6) holds.

Case 3. $v(0)+2=v(1)$. Since $v(1)\leqq 2$, we have $v(0)\leqq 0$. By Definition 5.1, $u(0)=u(1)$ or $u(1)-2$. By scaling, we may assume that $e_1=1$.

Let $u(0)=u(1)$. If $v(1)<2$, then $v(0)<0$. Thus, an argument similar to that of Case 1 shows that $L'_0\oplus H'_1$ and $M'_0\oplus H'_1$ contain the sublattice $\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ \pi^{v(0)} & 0\end{pmatrix}$. Hence, we may assume that $v(1)=2$ and $v(0)=0$. By Lemma 4.5, we see that $L'_0\oplus(\pi^{u(1)})\cong\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}\oplus\begin{pmatrix} & 1\\ 1 & D\end{pmatrix}\oplus(\pi^{u(1)})$, where $u(1)$ is odd. By condition 5) and the second argument of Case 2, we see that $L'_0\oplus H'_1$ and $M'_0\oplus H'_1$ both contain the sublattice $\begin{pmatrix} & 1\\ 1 & D\end{pmatrix}\oplus\begin{pmatrix} & 1\\ 0 & 0\end{pmatrix}$. Thus, A_6) holds in either case.

Let $u(0)=u(1)-2$. By Definition 5.1, we then have

$$e_0q_1(L)=\pi^2q_0(L)\subseteq q_1(L)=e_0{}^2q_1(L)\subseteq e_0q_1(L).$$

Thus, $q_0(L)=\pi^{-2}q_1(L)$ and we may assume that $e_0=1$ as well. If $v(1)<2$, then $u(0)<v(0)<0$, $-u(0)\geqq u(1)$ and $-v(0)\geqq v(1)$. Repeating the argument of Case 1, we see that $L'_0\oplus H'_1$ contains the sublattice

$$\begin{pmatrix} & 1\\ \pi^{u(0)} & \pi^{v(0)}b\end{pmatrix}\oplus\begin{pmatrix} & 1\\ \pi^{v(0)} & 0\end{pmatrix},\text{ where } b\in\Omega.$$

Similarly, $M'_0 \oplus H'_1$ contains the sublattice

$$\begin{pmatrix} & 1 \\ \pi^{u(0)} & \pi^{v(0)}b' \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ \pi^{v(0)} & 0 \end{pmatrix}, \text{ where } b' \in \Omega.$$

From $0 > u(0) + v(0) \equiv 1 \bmod 2$, condition 4) and Lemma 1.1, we see that $b = b'$. Finally, if $v(1) = 2$, then $u(0) < v(0) = 0$ and $-u(0) \geqq u(1)$. Repeating the argument of Case 1, we see that $L'_0 \oplus H'_1$ contains the sublattice $\begin{pmatrix} & 1 \\ \pi^{u(0)} & b \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ 1 & D \end{pmatrix}$ and that $M'_0 \oplus H'_1$ contains the sublattice $\begin{pmatrix} & 1 \\ \pi^{u(0)} & b' \end{pmatrix} \oplus \begin{pmatrix} & 1 \\ 1 & D' \end{pmatrix}$, where $b, b' \in \Omega$; $D, D' \in \{0, \lambda\}$; and $u(0) < 0$ is odd. By condition 5) and the second argument of Case 2, we get $D = D'$. By condition 4) and the last argument of the preceding paragraph, we see that $b = b'$. Thus A_6) holds.

From A_6), applying Lemma 1.3, we see that L''_0 and M''_0 are orthogonal summands of $L'_0 \oplus H'_1$ and $M'_0 \oplus H'_1$ respectively. Thus, we may "replace" L'_0 and M'_0 by L''_0 and M''_0 respectively. Hence, we have obtained decompositions of L and M satisfying A_1), A_2) as well as $L_0 \cong M_0$. Let L' and M' be the corresponding orthogonal complement of L_0 and M_0 in L and M. It is clear that L' and M' are normal lattices of type $(1, m)$ and satisfy a similar set of conditions as stated in our assertion. Hence, by induction, $L' \cong M'$. Thus, $L \cong M$. Q. E. D.

PRINCETON UNIVERSITY.

REFERENCES.

[1] C. Arf, "Untersuchungen über quadratische Formen in Körpern der Charakteristik 2," *Journal für die reine und angewandte Mathematik*, vol. 183 (1940), pp. 148-167.

[2] ———, "Über arithmetische Aequivalenz quadratische Formen in Potenzreihenkörpern über einem vollkommen Körpern der Charakterisk 2," *Revue de la Faculté des Sciences de l'Université d'Istanbul*, vol. 8 (1943), pp. 297-327.

[3] E. Artin, *Algebraic Numbers and Algebraic Functions*, New York University, 1951.

[4] ———, *Geometric Algebra*, New York, 1957.

[5] N. Bourbaki, "Algèbre," Chapitre 8, 9, *Actualités Scientifiques et Industrielles*, Nos. 1261, 1272, Paris, 1958-1959.

[6] P. Cartier, "On Arf and Dickson Invariants," (unpublished).

[7] C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable*, New York, 1951.

[8] J. Dieudonné, *Sur les groupes classiques*, Paris, 1948.

[9] M. Eichler, *Quadratische Formen und orthogonalen Gruppen*, Berlin, 1952.

[10] H. Hasse, "Über die Äquivalenz quadratische Formen im Körper der rationalen Zahlen," *Journal für die reine und angewandte Mathematik*, vol. 152 (1923), pp. 205-224.

[11] ———, "Äquivalenz quadratische Formen in einem beliebigen algebraischen Zahlkörper," *ibid.*, vol. 153 (1924), pp. 158-162.

[12] W. Klingenberg and E. Witt, "Über die Arfsche Invariante quadratische Formen mod 2," *ibid.*, vol. 193 (1954), pp. 121-122.

[13] H. Minkowski, *Gesammelte Abhandlungen I*, Leipzig, 1911.

[14] O. T. O'Meara, "Quadratic forms over local fields," *American Journal of Mathematics*, vol. 77 (1955), pp. 87-116.

[15] ———, "Integral equivalence of quadratic forms in ramified local fields," *ibid.*, vol. 79 (1957), pp. 157-186.

[16] E. Witt, "Theorie der quadratische Formen in beliebigen Körpern," *Journal für die reine und angewandte Mathematik*, vol. 176 (1937), pp. 31-44.

[17] ———, "Zyklische Körper und Algebren der Charakteristik p vom Grad p^n," *ibid.*, vol. 176 (1937), pp. 126-140.

[18] ———, "Über eine Invariante quadratische Formen mod 2," *ibid.*, vol. 193 (1954), pp. 119-120.

A GENERALIZED THEOREM OF KRULL-SEIDENBERG ON PARAMETRIZED ALGEBRAS OF FINITE TYPE.*

By Heisuke Hironaka.

1. **Introduction.** Let B be a *Noetherian ring* and A a *commutative B-algebra* such that $1_B 1_A = 1_A$, where 1_B and 1_A denote the unities of B and A respectively. We have a canonical homomorphism $i\colon B \to A$ such that $i(b)a = ba$ for $b \in B$ and $a \in A$. We shall say that $i\colon B \to A$ is *of finite type*, meaning that A is a commutative B-algebra of finite type, i. e., a homomorphic image of a polynomial ring over B. In general, if A' is a commutative ring with unity, $Spec(A')$ denotes the set of all prime ideal in A' which is furnished with *Zariski topology*; a subset X of $Spec(A')$ is, by defintion, open (respectively closed) if there exists an ideal $\mathfrak{A}$ in A' such that X is the set of those prime ideals in A' which do not (respectively do) contain $\mathfrak{A}$, where $\mathfrak{A}$ may be A' itself. The homomorphism $i\colon B \to A$ gives rise to a continuous mapping $f\colon Spec(A) \to Spec(B)$ defined by $f(\mathfrak{P}) = i^{-1}(\mathfrak{P})$ (= the prime ideal in B formed by those elements in B which are mapped into $\mathfrak{P}$ by i) for $\mathfrak{P} \in Spec(A)$. If $\mathfrak{P} \in Spec(A)$ and $\mathfrak{Q} = f(\mathfrak{P})$, there exists a unique homomorphism $i_{\mathfrak{P}}\colon B_{\mathfrak{Q}} \to A_{\mathfrak{P}}$ such that the following diagram is commutative:

$$\begin{array}{ccc} B & \xrightarrow{\;\; i \;\;} & A \\ \downarrow & & \downarrow \\ B_{\mathfrak{Q}} & \xrightarrow[\;\; i_{\mathfrak{P}} \;\;]{} & A_{\mathfrak{P}} \end{array}$$

where $B \to B_{\mathfrak{Q}}$ and $A \to A_{\mathfrak{P}}$ denote the canonical homomorphisms associated with the localizations. Similarly, if $\mathfrak{Q} \in Spec(B)$, there exists a homomorphism $i_{\mathfrak{Q}}\colon B_{\mathfrak{Q}} \to A_{\mathfrak{Q}}$, uniquely induced by i and the localizations, where $A_{\mathfrak{Q}}$ denotes the ring of quotients of A with respect to the multiplicatively closed subset $B - \mathfrak{Q}$ of B.

For our convenience, a Noetherian ring will be said to be *unmixed* if all the associated prime ideals of zero have the same corank (which is equal to the rank of the ring), and an ideal in it will be said to be *unmixed* if it has

* Received February 23, 1960.

no imbedded prime divisors. If A' is a Noetherian ring and $\mathfrak{A}'$ is an ideal in A', the unmixedness of $A'/\mathfrak{A}'$ is stronger than that of $\mathfrak{A}'$ itself. But, if A' satisfies the chain condition of prime ideals, an ideal $\mathfrak{A}'$ of rank r and generated by r elements ($r \geqq 0$) is unmixed if and only if $A'/\mathfrak{A}'$ is unmixed;[1] we know that if A' is an algebra of finite type over a field and is unmixed, then it satisfies the chain condition of prime ideals.

One of the theorems proved by W. Krull on the parameter-specialization of ideals in a polynomial ring may be stated as follows:

THEOREM OF KRULL.[2] *Let $B = k[u]$ with a field k and a variable u over k, and let us assume that $i\colon B \to A$ is of finite type. If B does not contain any zero-divisors of A and A_{B^*} is unmixed for $B^* = B - (0)$, then $A_{\mathfrak{Q}}/\mathfrak{Q}A_{\mathfrak{Q}}$ is unmixed and has the same rank as A_{B^*} (or, equivalently, the principal ideal $\mathfrak{Q}A_{\mathfrak{Q}}$ is unmixed and of rank one) for almost all (i.e., all but a finite number of) prime ideals $\mathfrak{Q}$ in B.*

A. Seidenberg generalized the theorem of Krull into the following form:

THEOREM OF SEIDENBERG.[3] *Let B be an arbitrary polynomial ring over a field, and let us assume that $i\colon B \to A$ is of finite type. If B does not contain any zero-divisors of A and A_{B^*} is unmixed, there exists a non-empty open subset U of $Spec(B)$ such that, for every $\mathfrak{Q} \in U$, $A_{\mathfrak{Q}}/\mathfrak{Q}A_{\mathfrak{Q}}$ is unmixed and has the same rank as A_{B^*}.*

The above condition on $\mathfrak{Q} \in U$ is equivalent to the following:[4]

---

[1] ***Krull's theorem: All the isolated prime divisors of an ideal with r generators in a Noetherian ring have ranks not greater than r.*** (Krull [2], and also M. Nagata, "Basic theorems on general commutative rings," *Memoirs of the college of science, Kyoto University*, vol. 24, No. 1 (1955)).

[2] Krull [3].

[3] Seidenberg [4], Appendix.

[4] Let k be the field of coefficients of B. Let K be the field of quotients of B. If $\mathfrak{p}$ is any of the associated prime ideals of zero in A, $\mathfrak{p}A_{B^*}$ is an associated prime ideal of zero in A_{B^*} because B does not contain any zero-divisor of A. Since A_{B^*} is unmixed, $\mathrm{corank}(\mathfrak{p}A_{B^*}) = \mathrm{rank}(A_{B^*})$. Noting that A is a k-algebra of finite type and A_{B^*} is a K-algebra of finite type, we get

$$\mathrm{corank}(\mathfrak{p}) = \dim_k(A/\mathfrak{p}) = \dim_k(B) + \dim_K(A_{B^*}) = \mathrm{rank}(B) + \mathrm{rank}(A_{B^*}).$$

Hence A is unmixed and $\mathrm{rank}(A) = \mathrm{rank}(B) + \mathrm{rank}(A_{B^*})$. Let $\mathfrak{Q}$ be any prime ideal in B such that $\mathfrak{Q}A_{\mathfrak{Q}} = A_{\mathfrak{Q}}$, and $\mathfrak{P}$ any prime ideal in A such that $\mathfrak{P}A_{\mathfrak{Q}}$ is a maximal ideal containing $\mathfrak{Q}A_{\mathfrak{Q}}$. Then $\dim_k(A/\mathfrak{P}) = \dim_k(B/\mathfrak{Q})$, hence, $\mathrm{corank}(\mathfrak{P}) = \mathrm{corank}(\mathfrak{Q})$. Therefore $\mathrm{rank}(A_{\mathfrak{P}}) = \mathrm{rank}(A) - \mathrm{corank}(\mathfrak{P}) = (\mathrm{rank}(B) + \mathrm{rank}(A_{B^*})) - \mathrm{corank}(\mathfrak{Q}) = \mathrm{rank}(\mathfrak{Q}) + \mathrm{rank}(A_{B^*})$. Now, $A_{\mathfrak{Q}}/\mathfrak{Q}A_{\mathfrak{Q}}$ is unmixed and of rank $= \mathrm{rank}(A_{B^*})$ if and only if $\mathfrak{Q}A_{\mathfrak{P}}$ $(= (x_1, \cdots, x_s)A_{\mathfrak{P}})$ is unmixed and of rank $= \mathrm{rank}(A_{\mathfrak{P}}) - \mathrm{rank}(A_{B^*})$ $(= s)$ for every $\mathfrak{P}$ as above if and only if $\mathfrak{Q}A_{\mathfrak{Q}}$ $(= (x_1, \cdots, x_s)A_{\mathfrak{Q}})$ is unmixed and of rank s.

(*) *if* $x_1,\cdots,x_s$ *is a regular system of parameters of* $B_{\mathfrak{Q}}$, *then* $(x_1,\cdots,x_s)A_{\mathfrak{Q}}$ *is unmixed and of rank* s,

and also to the following:

(**) *there exists a prime ideal* $\mathfrak{P}$ *in* A *such that* $\mathfrak{Q}=i^{-1}(\mathfrak{P})$, *and, for every such* $\mathfrak{P}$, $(x_1,\cdots,x_s)A_{\mathfrak{P}}$ *is unmixed and of rank* s *if* $x_1,\cdots,x_s$ *is a regular system of parameters of* $B_{\mathfrak{Q}}$, *where* $A_{\mathfrak{P}}$ *is view as a* $B_{\mathfrak{Q}}$-*algebra by means of* $i_{\mathfrak{P}}$.

We shall consider the following condition on $\mathfrak{P}\in Spec(A)$, which is stronger and more useful than the condition (**) viewed as a condition on a fixed $\mathfrak{P}$, in particular for $t=0$, that is:

$(***)_t$ *(for each non-negative integer* t*).*[5] *Let* $\mathfrak{Q}=i^{-1}(\mathfrak{P})$. *For every non-negative integer* $t'\leqq t$, *if* $x_1,\cdots,x_s$ *is any finite system of elements* $(s\geqq 0)$ *in* $B_{\mathfrak{Q}}$ *such that* $\operatorname{rank}(x_1,\cdots,x_s)B_{\mathfrak{Q}}=s$ *and if* $y_1,\cdots,y_{t'}$ *is any system of elements in* $A_{\mathfrak{P}}$ *such that* $\operatorname{rank}(x_1,\cdots,x_s,y_1,\cdots,y_{t'})A_{\mathfrak{P}}=\operatorname{rank}(x_1,\cdots,x_s)A_{\mathfrak{P}}+t'$, *then the ideal* $(x_1,\cdots,x_s,y_1,\cdots,y_{t'})A_{\mathfrak{P}}$ *is unmixed and of* rank $s+t'$.

In connection with this $(***)_t$, we shall also consider the following condition on a Noetherian ring A':

$(\text{—})_t$ *(for each non-negative integer* t*). For every non-negative integer* $t'\leqq t$, *if* $y_1,\cdots,y_{t'}$ *is any system of elements in* A' *such that*

$$\operatorname{rank}(y_1,\cdots,y_{t'})A'=t',$$

then the ideal $(y_1,\cdots,y_{t'})A'$ *is unmixed.*

In particular, the condition $(\text{—})_0$ is equivalent to saying that the zero ideal of A' is unmixed, and $(\text{—})_1$ is known as one of the important conditions for A' to be normal.

We state (and shall prove in § 2) the main result in this paper in the following form:

MAIN THEOREM. *(Generalized theorem of Krull-Seidenberg). Let us assume that* B *is a homomorphic image of a regular ring and that* $i\colon B\to A$ *is of finite type (in the sense explained in the very beginning). Then there*

[5] Throughout this paper, we adopt the following conventions: A finite system of elements, written as $x_1,\cdots,x_s$, denotes the empty system (i.e., the empty set) if $s=0$; the empty set of elements generates the zero ideal.

exists an open subset V_t of $Spec(A)$, for each non-negative integer t, such that $\mathfrak{P} \in V_t$ if and only if $\mathfrak{P}$ satisfies the condition $(***)_t$.

The homomorphism $i: B \rightarrow A$ induces a continuous mapping $f: Spec(A) \rightarrow Spec(B)$, which is called a *morphism* together with the family of homomorphisms $i_{\mathfrak{P}}: B_{f(\mathfrak{P})} \rightarrow A_{\mathfrak{P}}$ with $f(\mathfrak{P}) = i^{-1}(\mathfrak{P})$. One knows the following *theorem of C. Chevalley:*[6] *if $i: B \rightarrow A$ is of finite type, then the morphism $f: Spec(A) \rightarrow Spec(B)$ it induces maps every constructible subset of $Spec(A)$ onto a constructible subset of $Spec(B)$*; in general, a subset of a topological space is said to be *constructible* if it is a finite union of locally closed subsets, i. e., intersections of closed subsets and open subsets.

Let B and $i: B \rightarrow A$ be as in the main theorem. Let U'_t be the maximal open subset of $Spec(B)$ that is contained in the image $f(V_t)$ of the open subset V_t of $Spec(A)$ under the induced morphism $f: Spec(A) \rightarrow Spec(B)$. Let U''_t be the complement in $Spec(B)$ of the closure of the image $f(Spec(A) - V_t)$ of the closed complement of V_t in $Spec(A)$. We define the open subset V_t of $Spec(B)$ as the intersection of these U'_t and U''_t.

COROLLARY TO MAIN THEOREM. *If there exists an isolated prime ideal $\mathfrak{Q}'$ associated with zero in B such that $A_{\mathfrak{Q}'}$ satisfies the condition $(-)_t$, then the open subset U_t of $Spec(B)$, obtained as above, is not empty (in fact, contains $\mathfrak{Q}'$) and satisfies the following condition: For every $\mathfrak{Q} \in U_t$ there exists at least one $\mathfrak{P} \in Spec(A)$ such that $i^{-1}(\mathfrak{P}) = \mathfrak{Q}$, and if $\mathfrak{P} \in Spec(A)$ is such that $i^{-1}(\mathfrak{P}) \in U_t$, then $\mathfrak{P}$ satisfies the condition $(***)_t$; furthermore, if $\mathfrak{Q} \in U_t$ and if $x_1, \cdots, x_r$ is any system of parameters of $B_{\mathfrak{Q}}$, $A_{\mathfrak{Q}}/(x_1, \cdots, x_r)A_{\mathfrak{Q}}$ satisfies the condition $(-)_t$.*

The corollary is immediate from the main theorem and the definition of the open subset U_t; in fact, we can see that U_t is the maximal open subset of $Spec(B)$ which has the property stated in the corollary. In case $t = 0$, the corollary states a better generalization of the theorem of Krull than that of Seidenberg. The corollary for $t = 1$ is also proved by A. Seidenberg, under certin auxiliary conditions, in the course of his proof of the fact that almost all the hyperplane sections of a normal variety are also normal.[7]

We shall assume and make use of the whole statement of the following theorem due to M. Auslander:

---

[6] Chap. I, § 7, Grothendieck and Dieudonné [1]; Théorème 3. Also Cartan-Chevalley Seminar, 1955-56.

[7] See Seidenberg [4].

THEOREM 0. (*Theorem of Auslander*).[8] *Let A be a homomorphic image of a regular ring. Let $d(A_{\mathfrak{P}}) = \operatorname{rank}(A_{\mathfrak{P}}) - \operatorname{codh}(A_{\mathfrak{P}})$ for each $\mathfrak{P} \in Spec(A)$. Then, for each non-negative integer d, there exists a closed subset F_d of $Spec(A)$ such that $\mathfrak{P} \in F_d$ if and only if $d(A_{\mathfrak{P}}) \geqq d$.*

The above integer $\operatorname{codh}(A_{\mathfrak{P}})$ can be defined as follows: A system of elements in a local ring A', say $x_1, \cdots, x_s$, is defined as *a prime sequence of* A' (or, simply, *an A'-sequence*) if x_i is not a zero divisor in $A'/(x_1, \cdots, x_{i-1})A'$ for $s = 1, \cdots, s$, and if $(x_1, \cdots, x_s)A' \neq A'$ at the same time. An A'-sequence $x_1, \cdots, x_s$ is said to be *maximal* if $x_1, \cdots, x_s, y$ is not an A'-sequence for any element y in A'. Then one can prove:

(′) *all the maximal A'-sequence have the same length.*[9]

The integer $\operatorname{codh}(A')$ is, by definition, the length of any (hence, all the) maximal A'-sequences.

More generally, if S is any subset of a local ring A', a system of elements in S is called *an A'-sequence in S* if it is an A'-sequence. An A'-sequence in S, say $x_1, \cdots, x_s$, is said to be *maximal* if $x_1, \cdots, x_s, y$ is not an A'-sequence for any $y \in S$. As is easily seen, an A'-sequence in S, $x_1, \cdots, x_s$, is maximal if and only if at least one of the associated prime ideals of $(x_1, \cdots, x_s)A'$ contains $S \cap M'$, where M' denotes the maximal ideal of A'. Let $x_1, \cdots, x_s$ be any maximal A'-sequence in S. Let $\mathfrak{P}'$ be an associated prime ideal of $(x_1, \cdots, x_s)A'$ which contains $S \cap M'$. Then $x_1, \cdots, x_s$ is a maximal $A'_{\mathfrak{P}'}$-sequence and every A'-sequence in S is an $A'_{\mathfrak{P}'}$-sequence. Hence, by (′), s is the maximal integer attained by the lengths of A'-sequences in S. It follows that:

(″) *all the A'-sequences in S have the same length, for every subset S of A'.*

---

[8] This theorem was communicated by his letter to the author early in 1958. The letter included a complete and simple proof of the theorem, which was based on a characterization of $\operatorname{codh}(A_{\mathfrak{P}})$ (or, more generally, of homological codimension of an $A_{\mathfrak{P}}$-*module* of finite type) in terms of Ext-functors. His method als gives a proof of the following statement: *Let us assume that A is a homomorphic image of a regular ring and that all the maximal chains of prime ideals in the regular ring have the same length (for example, A is a k-algebra of finite type with a field k), then the set of those $\mathfrak{P}$ such that* $\operatorname{corank}(\mathfrak{P}) + \operatorname{codh}(A_{\mathfrak{P}}) \leqq d$ *is a closed subset of $Spec(A)$ for each non-negative integer d.* One can see that this conclusion is stronger than that of Theorem 0. Consequently, Theorem 1 can be restated and proved likewise.

For the proof of Theorem 0, a paper by M. Auslander (forthcoming in *Illinois Journal of Mathematics*).

[9] Prop. 3.4, Auslander and Buchsbaum [5]; also [6].

We define $\operatorname{codh}(A', S)$ as the length of any (hence, all the) maximal A'-sequences in S. In this paper we shall freely use the following convention: For an arbitrary set S (not necessarily a subset of A'), if there is given a mapping $i \colon S \to A'$ (or, every element of S operates on A' like multiplication by a certain element in A') and if there occurs no ambiguity about it by the context, then we write $\operatorname{codh}(A', S)$ instead of $\operatorname{codh}(A', i(S))$.[10]

We shall first prove the following generalization of the Theorem 0 (See § 2):

THEOREM 1. *Let us assume that B is a homomorphic image of a regular ring and that $i \colon B \to A$ is of finite type. Then, for each non-negative integer d, there exists a closed subset G_d of $Spec(A)$ such that $\mathfrak{P} \in G_d$ if and only if $d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) \geqq d$ with $\mathfrak{Q} = i^{-1}(\mathfrak{P})$, where*

$$d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = \operatorname{rank}(B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$$

by definition.

We shall see that the integer $d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$ is always non-negative (Lemma 2 in § 2). Therefore *the set of those $\mathfrak{P} \in Spec(A)$ for which $d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0$ is an open subset of $Spec(A)$*. This is the only part of Theorem 1 that we shall use later (in the proof of Main Theorem). We shall see in § 2 that Main Theorem is an immediate consequence of this result and Theorem 0.

Previously $(***)_t$ for non-negative integers t are introduced as conditions on $\mathfrak{P} \in Spec(A)$ for the given B and B-algebra A. More generally, however, they may be viewed as conditions on a given pair of a local ring B' and a local B'-algebra A' (in the places of $B_{\mathfrak{Q}}$ and $A_{\mathfrak{P}}$ respectively).

As will be seen by Lemma 4 of § 2, the condition: $d(A', B') = 0$ for a given pair of a local ring B' and a local B'-algebra A' is equivalent to saying that

$(***)_{-1}$ *every finite system of elements of B', say s elements, which generates an ideal of rank s in B', is an A'-sequence.*

Furthermore, for the sake of convenience, as $(***)_{-d}$ will be regarded the condition: $d(A', B') < d$, for each positive integer d. Theorem 1 will then be considered as a part of Main Theorem by saying that:

Main Theorem is true for all integers t, non-negative and negative, and $(***)_t$ is stronger than $(***)_{t'}$ if $t > t'$; consequently, the corollary to Main Theorem is also true for all integers t (where the condition $(-)_t$ with $t < 0$ is empty).

[10] An analogous convention will be used in saying that a system of elements in S is (or is not) an A'-sequence.

In § 3, we shall prove the following proposition, which is not new:

PROPOSITION. *Let B' be a local ring and A' a local B'-algebra. Let $\mathfrak{J}'$ be an ideal in A' such that B' and the B'-algebra $A'/\mathfrak{J}'$ satisfy the condition $(***)_{-1}$. Let $x_1, \cdots, x_s$ be any system of parameters of B' and set $\bar{A}' = A'/(x_1, \cdots, x_s)A'$. Then, for an ideal $\mathfrak{J}'_0$ in A', $\mathfrak{J}'_0 \subseteq \mathfrak{J}'$ and $\mathfrak{J}'_0\bar{A}' = \mathfrak{J}'\bar{A}'$ imply $\mathfrak{J}'_0 = \mathfrak{J}'$.*

Main Theorem, in virtue of this proposition, can be proved also useful to study the property of complete intersection (local or even global) of cycles in an algebraic family. As for cycles [11] of codimension one, for example, we shall prove the following

THEOREM 2. *Let V be a normal variety and $\{X_t\}_{t \in T}$ an algebraic family of V-cycles of codimension one with a non-singular parameter variety T. Then there exists a non-empty open subset T' of T (in Zariski topology) such that either all or none of the V-cycles X_t with $t \in T'$ are locally principal (everywhere) on V.*

This theorem will be proved as an application of the main theorem (or rather, of its corollary) for $t = 0$. (Some different proofs of this theorem have been suggested to the author by A. Grothendieck and also by C. S. Seshadri.)

Theorem 2 will be used in a forthcoming paper concerning the notion of algebraic equivalence and the theory of blowing-up (factorization of birational morphisms); in the paper we shall define a certain algebraic equivalence of *Ideals* (= coherent algebraic sheaves of fractional ideals) on an algebraic model. We shall also see certain relations between the two notions of algebraic equivalences, the one defined for arbitrary cycles of codimension one (in the sense of A. Weil) [10] and the other for *divisors in Cartier's sense* (= locally principal cycles of codim. 1) introduced by C. Chevalley.[12] For example, the following results will be included: (In the terminology of Cartier-Chevalley)

((1)) *Let V be a normal variety (defined over a fixed algebraically closed field). Let $G(V)$ denote the additive group of divisors on V, and $G_a(V)$ the subgroup of $G(V)$ of those which are algebraically equivalent to*

[11] See Weil [7], comments on equivalence relations of algebraic cycles. Also, to study the Picard variety of a normal variety from this point of view, S. Lang's book, *Abelian Varieties* (Interscience, New York (No. 7)), is complete.

[12] See a paper of C. Chevalley, "On the Picard variety of a normal variety," forthcoming in the *American Journal of Mathematics* (1959-60).

zero on V. Let V'' be an open subset of V such that $V - V''$ has codimension at least two. Then we have a monomorphism r of $G(V)$ into $G(V'')$ (restriction of divisors to V'') which have the following properties: $G_a(V) \subseteq r^{-1}(G_a(V''))$ (—clear—) and the quotient $r^{-1}(G_a(V''))/G_a(V)$ is a finite group.

((2)) *Let us consider the pairs (V',f) of a normal variety V' and a porper birational morphism $f\colon V' \to V$ (i.e., a birational transformation of V' onto V which is regular everywhere on V' and such that V' is complete over V with respect to f). If $\mathfrak{J}$ is a coherent algebraic sheaf of fractional ideals on V, $f^*(\mathfrak{J})$ shall denote the coherent algebraic sheaf of fractional ideals on V' generated by the ideals of $\mathfrak{J}$, where the function fields of V and V' are canonically identified by means of f. Let V'' denote the open subset of V which consists of all the simple points of V. Then, given a coherent sheaf $\mathfrak{J}$ of fractional ideals on V, there exists a pair (V',f) such that $f^*(\mathfrak{J})$ is locally principal on V' and is associated with a divisor which is algebraically equivalent to zero on V', if and only if $\mathfrak{J}$ itself is locally principal on V and is associated with a divisor belonging to $r^{-1}(G_a(V''))$.*

((3)) *Let V and (V',f) be as above. Let V'' denote the open subset of V' (also, of V) which consists of those points of V' at which f is biregular. Let r' denote the homomorphism of $G(V')$ into $G(V'')$ (restriction of divisors to V''). Then the homomorphism $G_a(V')/G_l(V') \to G_a(V'')/G_l(V'')$ is injective, where $G_l(V')$ (resp. $G_l(V'')$) denotes the subgroup of principal (linearly equivalent to zero) divisors on V' (resp. V'').*

These theorems (especially ((2))) will play important roles in the study on the lattice of normal projective factors of a birational morphism.

2. Proofs of Theorem 1 and Main Theorem. Let B' be a local ring and A' a local B'-algebra. We shall consider the following condition on A' over B':

(!) *For every ideal $\mathfrak{B}$ in B',* $\operatorname{rank}(\mathfrak{B}) = \operatorname{rank}(\mathfrak{B}A')$.

It is necessary for A to satisfy the condition (!) over B' that A' *dominates* B', i.e., the maximal ideal of B' is the complete inverse image of the maximal ideal of A' under the canonical homomorphism of B' into A'. One can easily prove that if A' is a local ring of an integral extension of a polynomial ring over B', and if A' dominates B', then A' satisfies the condition (!) over B'.

LEMMA 1. *Let B be a Noetherian ring and let $i\colon B \to A$ be of finite*

type. Then there can be found a finite set of prime ideals, $\mathfrak{P}^j$, in A which possesses the following property: for every $\mathfrak{P} \in Spec(A)$, there exists at least one $\mathfrak{P}^j$ such that $\mathfrak{P}^j \subseteq \mathfrak{P}$ and $A_{\mathfrak{P}}/\mathfrak{P}^j A_{\mathfrak{P}}$ satisfies the condition (!) over $B_{\mathfrak{Q}}/\mathfrak{Q}^j B_{\mathfrak{Q}}$, where $\mathfrak{Q} = i^{-1}(\mathfrak{P})$ and $\mathfrak{Q}^j = i^{-1}(\mathfrak{P}^j)$.

Proof. The proof is by induction on $n = \mathrm{rank}(A)$. If $\mathrm{rank}(A) = 0$, then A has only a finite number of prime ideals and there is nothing to prove. Let us assume that $\mathrm{rank}(A) = n > 0$ and that the lemma is true for $\mathrm{rank}(A) < n$. First take all the isolated prime ideals belonging to zero in A, say $\mathfrak{P}_0{}^j$. For each j, set $A/\mathfrak{P}_0{}^j = A^j$ and $B/\mathfrak{Q}_0{}^j = B^j$, where $\mathfrak{Q}_0{}^j = i^{-1}(\mathfrak{P}_0{}^j)$. Then the homomorphism i induces $i^j\colon B^j \to A^j$, which is injective and of finite type. By the normalization theorem, there exists a non-zero element b_j in B^j such that $A^j[b_j{}^{-1}]$ is integral over a polynomial ring over $B^j[b_j{}^{-1}]$. Take then all the isolated prime ideals belonging to $b_jA + \mathfrak{P}_0{}^j$, say $\mathfrak{P}^{jk}$. By the induction assumption, there can be found a finite set of prime ideals $\bar{\mathfrak{P}}^{jkm}$ in $A/\mathfrak{P}^{jk}$, for each $\mathfrak{P}^{jk}$, which possesses the property, stated in the lemma, for $A/\mathfrak{P}^{jk}$ viewed as B-algebra in a natural way. Let $\mathfrak{P}^{jkm}$ be the prime ideal in A such that $\mathfrak{P}^{jkm} \supset \mathfrak{P}^{jk}$ and $\mathfrak{P}^{jkm}/\mathfrak{P}^{jk} = \bar{\mathfrak{P}}^{jkm}$. Then the finite set $\{\mathfrak{P}_0{}^j, \mathfrak{P}^{jk}, \mathfrak{P}^{jkm}\}$ possesses the property for A as B-algebra.

LEMMA 2. *Let B and $i\colon B \to A$ be as in Lemma 1. If $\mathfrak{P}$ and $\mathfrak{P}'$, $\in Spec(A)$, are such that $\mathfrak{P} \supseteq \mathfrak{P}'$, then always*

$$d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) \geqq d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \geqq 0,$$

where $\mathfrak{Q} = i^{-1}(\mathfrak{P})$ and $\mathfrak{Q}' = i^{-1}(\mathfrak{P}')$.

Proof. It is clear from the definition of rank and corank that

$$(1) \quad \mathrm{rank}(B_{\mathfrak{Q}}) - \mathrm{rank}(B_{\mathfrak{Q}'}) \geqq \mathrm{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}).$$

Therefore, to prove the first inequality in the lemma, it is sufficient to prove

$$(2) \quad \mathrm{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) - \mathrm{codh}\,(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \leqq \mathrm{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}),$$

where $\mathrm{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'})$ denotes $\mathrm{codh}(A_{\mathfrak{P}'}, i_{\mathfrak{P}'}(B_{\mathfrak{Q}'}))$ and likewise $\mathrm{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$. We can prove (2) by induction on $\mathrm{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) = s$. If $s = 0$ then $\mathfrak{Q} = \mathfrak{Q}'$, $B_{\mathfrak{Q}} = B_{\mathfrak{Q}'}$ and $A_{\mathfrak{P}'} = (A_{\mathfrak{P}})_{\mathfrak{P}'A_{\mathfrak{P}}}$ (the ring of quotients of $A_{\mathfrak{P}}$ with respect to the prime ideal $\mathfrak{P}'A_{\mathfrak{P}}$). Hence, every $A_{\mathfrak{P}}$-sequence in $B_{\mathfrak{Q}}$ is an $A_{\mathfrak{P}'}$-sequence in $B_{\mathfrak{Q}'}$, which shows $\mathrm{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) \leqq \mathrm{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'})$ and the inequality (2) for $s = 0$. Let us assume that $s > 0$ and that (2) is verified for $\mathrm{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) < s$. To prove the inequality for $\mathfrak{P}$, $\mathfrak{P}'$ with $\mathrm{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) = s$, let us take a maximal $A_{\mathfrak{P}}$-sequence in $\mathfrak{Q}'B_{\mathfrak{Q}}$, say $x_1, \cdots, x_t$. As is easily seen, $x_1, \cdots, x_t$ is an $A_{\mathfrak{P}'}$-sequence and its image in $B_{\mathfrak{Q}'}$ is an $A_{\mathfrak{P}'}$-sequence in $B_{\mathfrak{Q}'}$. Therefore

(3) $\operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \geqq t.$

If $x_1, \cdots, x_t$ is a maximal $A_{\mathfrak{P}}$-sequence in $B_{\mathfrak{Q}}$, i.e., $\operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = t$, then the left hand side of (2) is not positive by (3) and the inequality (2) is obvious. Hence we may proceed to the case in which we have an element x_{t+1} in $B_{\mathfrak{Q}}$ such that $x_1, \cdots, x_t, x_{t+1}$ is an $A_{\mathfrak{P}}$-sequence in $B_{\mathfrak{Q}}$. There exists an associated prime ideal of $(x_1, \cdots, x_t, x_{t+1})A_{\mathfrak{P}}$ which contains $\mathfrak{Q}'B_{\mathfrak{Q}}$, say $\mathfrak{P}''A_{\mathfrak{P}}$ with $\mathfrak{P}'' \in Spec(A)$. In fact, if it were not the case, there would be an element y in $\mathfrak{Q}'B_{\mathfrak{Q}}$ such that $x_1, \cdots, x_t, x_{t+1}, y$ is an $A_{\mathfrak{P}}$-sequence; so that $x_1, \cdots, x_t, y$ would be an $A_{\mathfrak{P}}$-sequence in $\mathfrak{Q}'B_{\mathfrak{Q}}$, in view of the following result:

$(')_1$ *In a local ring, a finite number of elements form a prime sequence in every ordering if they do in any ordering.*[13]

On the other hand there should not be such y by the selection of $x_1, \cdots, x_t$, and we can conclude the existence of such $\mathfrak{P}''$. Now, let $\mathfrak{Q}'' = i^{-1}(\mathfrak{P}'')$. $\mathfrak{Q}'' \subseteqq \mathfrak{Q}$ and $\mathfrak{Q}''B_{\mathfrak{Q}} \supseteqq (\mathfrak{Q}', x_{t+1})B_{\mathfrak{Q}}$. But x_{t+1} is not in $\mathfrak{Q}'B_{\mathfrak{Q}}$ because $x_1, \cdots, x_t$ is maximal as an $A_{\mathfrak{P}}$-sequence *in* $\mathfrak{Q}'B_{\mathfrak{Q}}$ and, at the same time $x_1, \cdots, x_t, x_{t+1}$ is an $A_{\mathfrak{P}}$-sequence. It follows that $\operatorname{corank}(\mathfrak{Q}''B_{\mathfrak{Q}}) \leqq \operatorname{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) - 1$. So, by the induction assumption applied to $\mathfrak{P}$ and $\mathfrak{P}''$, we get

(4) $\operatorname{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) - 1 \geqq \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}''}, B_{\mathfrak{Q}''}),$

where $\mathfrak{Q}'' = i^{-1}(\mathfrak{P}'')$.

Since $\mathfrak{P}''A_{\mathfrak{P}}$ is an associated prime ideol of $(x_1, \cdots, x_t, x_{t+1})A_{\mathfrak{P}}$, $x_1, \cdots, x_t, x_{t+1}$ is a maximal $A_{\mathfrak{P}''}$-sequence and, of course, a maximal $A_{\mathfrak{P}''}$-sequence in $B_{\mathfrak{Q}''}$. It shows that

(5) $\operatorname{codh}(A_{\mathfrak{P}''}, B_{\mathfrak{Q}''}) = t + 1.$

The inequality (2) is now immediate as follows:

$$\begin{aligned} \operatorname{corank}(\mathfrak{Q}'B_{\mathfrak{Q}}) &\geqq \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}''}, B_{\mathfrak{Q}''}) + 1 && \text{(by (4))} \\ &= \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) - t && \text{(by (5))} \\ &\geqq \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) && \text{(by (3)).} \end{aligned}$$

The first inequality is now proved. The second one is an immediate consequence of the first. In fact, if $\mathfrak{P}^*$ is an associated prime ideal of zero in A which is contained in $\mathfrak{P}'$, then

$$d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \geqq d(A_{\mathfrak{P}^*}, B_{\mathfrak{Q}^*}) - \operatorname{rank}(B_{\mathfrak{Q}^*}) \geqq 0,$$

where $\mathfrak{Q}^* = i^{-1}(\mathfrak{P}^*)$.

[13] Cor. 2.9 of Prop. 2.8, Auslander and Buchsbaum [6].

We shall consider a chain condition of the following type, on the prime ideals in a Noetherian ring B':

• (*The restricted chain condition*). *For every pair of prime ideals $\mathfrak{Q}$ and $\mathfrak{Q}'$ in B' such that $\mathfrak{Q} \subseteq \mathfrak{Q}'$, all the maximal chains of strictly ascending prime ideals between $\mathfrak{Q}$ and $\mathfrak{Q}'$ have the same length.*

It is clear that if the condition is satisfied in B', then so in every homomorphic image of B', and that, for example, it is satisfied in every homomorphic image of a regular ring.

LEMMA 3. *Let us assume that B satisfies the restricted chain condition on the prime ideals (in the above sense), that $i\colon B \rightarrow A$ is of finite type, and that A is a homomorphic image of a regular ring. Let $\mathfrak{P}$ be any prime ideal in A and $\mathfrak{Q} = i^{-1}(\mathfrak{P})$. Then there exists an open subset V of $Spec(A)$ such that $\mathfrak{P} \in V$ and $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$ for every $\mathfrak{P} \subset \mathfrak{P}' \in V$ and $\mathfrak{Q}' = i^{-1}(\mathfrak{P}')$.*

Proof. First we shall reduce the proof to the case in which $\operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0$. To do it, we take a maximal $A_{\mathfrak{P}}$-sequence in B (which is necessarily such one in $B_{\mathfrak{Q}}$), say $x_1, \cdots, x_q$ with $q = \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$. Replacing A by $A[u^{-1}]$ with a suitable $u \in A - \mathfrak{P}$, we may assume that $x_1, \cdots, x_q$ is an A-sequence in B. If the lemma is true in case $\operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0$, there exists an open subset $\bar{V}$ of $Spec(\bar{A})$, $\bar{A}$ denoting $A/(x_1, \cdots, x_q)A$ which is viewed as a B-algebra in the natural way, such that $\bar{\mathfrak{P}} = \mathfrak{P}/(x_1, \cdots, x_q)A$ is in $\bar{V}$ and $d(\bar{A}_{\bar{\mathfrak{P}}'}, B_{\mathfrak{Q}'}) = d(\bar{A}_{\bar{\mathfrak{P}}}, B_{\mathfrak{Q}})$ for every $\bar{\mathfrak{P}} \subseteq \bar{\mathfrak{P}}' \in \bar{V}$ and $\bar{\mathfrak{P}}' = \mathfrak{P}'/(x_1, \cdots, x_q)A$. Take an open subset V of $Spec(A)$ which induces the open subset $\bar{V}$ in $Spec(\bar{A})$; for example, the subset consisting of those $\mathfrak{P}' \in Spec(A)$ such that either $\mathfrak{P}' \not\supseteq (x_1, \cdots, x_q)A$ or $\mathfrak{P}'/(x_1, \cdots, x_q)A = \bar{\mathfrak{P}}' \in \bar{V}$. This open subset V of $Spec(A)$ satisfies the condition stated in the lemma. In fact, if $\mathfrak{P} \subset \mathfrak{P}' \in V$, then $\bar{\mathfrak{P}} \subset \bar{\mathfrak{P}}' \in \bar{V}$ and therefore

$$\begin{aligned} d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) &= \operatorname{rank}(B_{\mathfrak{Q}'}) - \operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \\ &= \operatorname{rank}(B_{\mathfrak{Q}'}) - (\operatorname{codh}(\bar{A}_{\bar{\mathfrak{P}}'}, B_{\mathfrak{Q}'}) + q) = d(\bar{A}_{\bar{\mathfrak{P}}'}, B_{\mathfrak{Q}'}) - q \\ &= d(\bar{A}_{\bar{\mathfrak{P}}}, B_{\mathfrak{Q}}) - q = \operatorname{rank}(B_{\mathfrak{Q}}) - (\operatorname{codh}(\bar{A}_{\bar{\mathfrak{P}}}, B_{\mathfrak{Q}}) + q) \\ &= \operatorname{rank}(B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}). \end{aligned}$$

The reduction of the proof is now done, and from now on we assume that $\operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0$. First, let us take an element $u \in B - \mathfrak{Q}$ which is contained in all but those contained in $\mathfrak{Q}$ of the associated prime ideals of zero in B. Let V' be the open subset of $Spec(A)$ which is defined by the ideal uA.

Then

(6) $\mathfrak{P} \in V'$, *and* $\mathrm{rank}(B_{\mathfrak{Q}'}) - \mathrm{rank}(B_{\mathfrak{Q}}) = \mathrm{corank}(\mathfrak{Q}B_{\mathfrak{Q}'})$ *if* $\mathfrak{P} \subset \mathfrak{P}' \in V'$ *and* $\mathfrak{Q}' = i^{-1}(\mathfrak{P}')$.

The equality in (6) follows from the assumption that B satisfies the restricted chain condition on the prime ideals, because $\mathfrak{Q}$ contains all of those associated prime ideals of zero in B which are contained in $\mathfrak{Q}'$. Next, we apply Theorem 0 in § 1 to A. Let F_d be the closed subsets of $Spec(A)$ and take the isolated prime ideals $\mathfrak{P}_{dk}$ of F_d for every d. We have only a finite number of such $\mathfrak{P}_{dk}$ because F_d are empty for all $d > \mathrm{rank}(A)$. Then we apply Lemma 1 to the B-algebra $A/\mathfrak{P}_{dk}$, for each $\mathfrak{P}_{dk}$, and choose a finite set of prime ideals $\mathfrak{P}_{dk}{}^j$ in A such that all the $\mathfrak{P}_{dk}{}^j$ contain $\mathfrak{P}_{dk}$ and, if $\mathfrak{P}_{dk} \subseteq \mathfrak{P}'' \in Spec(A)$, $A_{\mathfrak{P}''}/\mathfrak{P}_{dk}{}^j A_{\mathfrak{P}''}$ satisfies the condition (!) over $B_{\mathfrak{Q}''}/\mathfrak{Q}_{dk}{}^j B_{\mathfrak{Q}''}$ for at least one one $\mathfrak{P}_{dk}{}^j$, where $\mathfrak{Q}_{dk}{}^j = i^{-1}(\mathfrak{P}_{dk}{}^j)$. When $\mathfrak{P}'' \in Spec(A)$ is given, we can choose $\mathfrak{P}_{dk}$ such that $\mathfrak{P}_{dk} \subseteq \mathfrak{P}''$ and $d(A_{\mathfrak{P}''}) = d(A_{\mathfrak{P}_{dk}})$, hence $d(A_{\mathfrak{P}''}) = d(A_{\mathfrak{P}_{dk}{}^j})$ for every $\mathfrak{P}_{dk}{}^j \subseteq \mathfrak{P}''$. Therefore, denoting by $\mathfrak{P}^*$ any one of the $\mathfrak{P}_{dk}{}^j$ (all d, k and j), we can state:

(7) *There exists a finite set of prime ideals* $\mathfrak{P}^*$ *in* A *which has the following property: for every* $\mathfrak{P}'' \in Spec(A)$, *there can be found at least one* $\mathfrak{P}^*$ *such that* $\mathfrak{P}'' \supseteq \mathfrak{P}^*$, $d(A_{\mathfrak{P}''}) = d(A_{\mathfrak{P}^*})$ *and* $A_{\mathfrak{P}''}/\mathfrak{P}^* A_{\mathfrak{P}''}$ *satisfies the condition* (!) *over* $B_{\mathfrak{Q}''}/\mathfrak{Q}^* B_{\mathfrak{Q}''}$ *with* $\mathfrak{Q}^* = i^{-1}(\mathfrak{P}^*)$ *and* $\mathfrak{Q}'' = i^{-1}(\mathfrak{P}'')$.

Let V'' be the intersection of those open subsets of $Spec(A)$ which are defined by the ideals $\mathfrak{P}^* \not\subseteq \mathfrak{P}$. Since we have only a finite number of $\mathfrak{P}^*$, V'' is an open subset of $Spec(A)$. Clearly $\mathfrak{P} \in V''$.

(8) *If* $\mathfrak{P}'' \in V''$, *then* $\mathfrak{P}'' \supseteq \mathfrak{P}^*$ *implies* $\mathfrak{P} \supseteq \mathfrak{P}^*$.

We can now prove that the open subset $V = V' \cap V''$ of $Spec(A)$, containing $\mathfrak{P}$, possesses the property stated in the lemma. Let $\mathfrak{P} \subset \mathfrak{P}' \in V$ and $\mathfrak{Q}' = i^{-1}(\mathfrak{P}')$. We want to prove that $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$. Let us take a maximal $A_{\mathfrak{P}'}$-sequence in $B_{\mathfrak{Q}'}$, $x_1, \cdots, x_s$. By the maximality, there exists an associated prime ideal of $(x_1, \cdots, x_s)A_{\mathfrak{P}'}$ which contains the maximal ideal of $B_{\mathfrak{Q}'}$, say $\mathfrak{P}''A_{\mathfrak{P}'}$ with $\mathfrak{P}'' \in Spec(A)$. Then $\mathfrak{Q}'' = i^{-1}(\mathfrak{P}'') = \mathfrak{Q}'$, and $x_1, \cdots, x_s$ is a maximal $A_{\mathfrak{P}''}$-sequence, hence,

(9) $\mathrm{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = s = \mathrm{codh}(A_{\mathfrak{P}''})$.

Let us take and fix one $\mathfrak{P}^*$, for $\mathfrak{P}''$, having the property stated in (7). Since $\mathfrak{P}'$ is in V'', $\mathfrak{P}''$ is also in V''. By (8), $\mathfrak{P}^* \subseteq \mathfrak{P}$. Therefore if $\mathfrak{Q}^* = i^{-1}(\mathfrak{P}^*)$, $\mathfrak{Q}^* \subseteq \mathfrak{Q}$ and $\mathrm{corank}(\mathfrak{Q}^* B_{\mathfrak{Q}''}) = \mathrm{corank}(\mathfrak{Q}^* B_{\mathfrak{Q}'}) \geqq \mathrm{corank}(\mathfrak{Q}B_{\mathfrak{Q}'})$. By (7),

$d(A_{\mathfrak{P}''}) = d(A_{\mathfrak{P}^*})$, i. e., $\operatorname{rank}(A_{\mathfrak{P}'}) - \operatorname{codh}(A_{\mathfrak{P}''}) = \operatorname{rank}(A_{\mathfrak{P}^*}) - \operatorname{codh}(A_{\mathfrak{P}^*})$. Hence

$$\begin{aligned} \operatorname{codh}(A_{\mathfrak{P}''}) &\geqq \operatorname{codh}(A_{\mathfrak{P}''}) - \operatorname{codh}(A_{\mathfrak{P}^*}) \\ &= \operatorname{rank}(A_{\mathfrak{P}''}) - \operatorname{rank}(A_{\mathfrak{P}^*}) \geqq \operatorname{corank}(\mathfrak{P}^* A_{\mathfrak{P}''}). \end{aligned}$$

Again by (7) $A_{\mathfrak{P}''}/\mathfrak{P}^* A_{\mathfrak{P}''}$ satisfies the condition (!) over $B_{\mathfrak{Q}''}/\mathfrak{Q}^* B_{\mathfrak{Q}''}$ (See the paragraph preceding Lemma 1), hence,

$$\operatorname{rank}(A_{\mathfrak{P}''}/\mathfrak{P}^* A_{\mathfrak{P}''}) \geqq \operatorname{rank}(B_{\mathfrak{Q}''}/\mathfrak{Q}^* B_{\mathfrak{Q}''}),$$

i. e., $\operatorname{corank}(\mathfrak{P}^* A_{\mathfrak{P}''}) \geqq \operatorname{corank}(\mathfrak{Q}^* B_{\mathfrak{Q}''})$. We have seen that $\operatorname{corank}(\mathfrak{Q}^* B_{\mathfrak{Q}''})$ $\operatorname{corank}(\mathfrak{Q} B_{\mathfrak{Q}'})$ and that $\operatorname{codh}(A_{\mathfrak{P}''}) \geqq \operatorname{corank}(\mathfrak{P}^* A_{\mathfrak{P}''})$. We therefore get (by these three inequalities)

$$\text{(10)} \quad \operatorname{codh}(A_{\mathfrak{P}''}) \geqq \operatorname{corank}(\mathfrak{Q} B_{\mathfrak{Q}'}).$$

Thus $\operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) - \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = \operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'})$

$$\begin{aligned} &(\text{because } \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0 \text{ by assumption}) \\ &\qquad = \operatorname{codh}(A_{\mathfrak{P}''}) && (\text{by } (9)) \\ &\qquad \geqq \operatorname{corank}(\mathfrak{Q} B_{\mathfrak{Q}'}) && (\text{by } (10)) \\ &\qquad = \operatorname{rank}(B_{\mathfrak{Q}'}) - \operatorname{rank}(B_{\mathfrak{Q}}). \end{aligned}$$

The last equality is by (6); we took $\mathfrak{P}'$ such that $\mathfrak{P} \subseteq \mathfrak{P}' \in V \subseteq V'$. The result can be written as

$$\operatorname{rank}(B_{\mathfrak{Q}'}) - \operatorname{codh}(A_{\mathfrak{P}'}, B_{\mathfrak{P}'}) \leqq \operatorname{rank}(B_{\mathfrak{Q}}) - \operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$$

or, $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) \leqq d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$. By Lemma 2 we can conclude the equality $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$.[14]

THEOREM* 1.[15] *Let us assume that B satifies the restricted chain condition on the prime ideals, that $i\colon B \to A$ is of finite type and that A is a homomorphic image of a regular ring. Then, for each non-negative integer d, there exists a closed subset G_d of Spec(A) such that $\mathfrak{P} \in G_d$ if and only if $d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) \geqq d$ with $\mathfrak{Q} = i^{-1}(\mathfrak{P})$.*

---

[14] Notice that the inclusion relation of $\mathfrak{P}$ and $\mathfrak{P}'$ (and consequently the inequality in the conclusion) was reversed in Lemma 2.

[15] The assumption of Theorem* 1 is weaker than that of Theorem 1 announced in the introduction (§ 1). In fact, if B is a homomorphic image of a regular ring, then it satisfies the restricted chain condition, and, moreover, if A is a B-algebra of finite type, then A is also a homomorphic image of a regular ring, because a polynomial ring over a regular ring is also regular. (Furthermore, it can be seen that Theorem* 1 remains true even without the restricted chain condition in B.)

Proof. Let us fix a non-negative integer d. Let us denote by V_d the set of those $\mathfrak{P} \in Spec(A)$ such that $d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) < d$ with $\mathfrak{Q} = i^{-1}(\mathfrak{P})$. The theorem amounts to saying that V_d is an open subset of $Spec(A)$. Denoting by V_d' *the maximal* open subset contained in V_d (which exists because an arbitrary union of open subsets is open), we shall prove that $V_d' = V_d$. Suppose $V_d' \neq V_d$. Let us take an *isolated* prime ideal $\mathfrak{P}$ in $V_d - V_d'$. First, we see that $\mathfrak{P}$ is isolated even in $Spec(A) - V_d'$. In fact, if $\mathfrak{P} \supseteq \mathfrak{P}' \in Spec(A) - V_d'$, then, by Lemma 2, $\mathfrak{P}' \in V_d$, hence, $\in V_d - V_d'$. Therefore $\mathfrak{P}$, isolated in $V_d - V_d'$ coincides with $\mathfrak{P}'$, i.e., $\mathfrak{P}$ is isolated in $Spec(A) - V_d'$. Since V_d' is open, there exist only a finite number of isolated prime ideals in $Spec(A) - V_d'$, say $\mathfrak{P}_1 = \mathfrak{P}, \mathfrak{P}_2, \cdots, \mathfrak{P}_p$. Let V' be the open subset of $Spec(A)$ defined by the ideal $\mathfrak{P}_2 \cap \cdots \cap \mathfrak{P}_p$ (or, the whole $Spec(A)$ if $p = 1$). Clearly $\mathfrak{P} \in V'$. On the other hand, there exists an open subset V of $Spec(A)$ such that $\mathfrak{P} \in V$ and $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}})$ if $\mathfrak{P} \subset \mathfrak{P}' \in V$. (See Lemma 3). Let $V'' = V \cap V'$. Then V'' is an open subset of $Spec(A)$ containing $\mathfrak{P}$. What is more, V'' is contained in V_d. In fact, let $\mathfrak{P}' \in V''$. If $\mathfrak{P}' \supseteq \mathfrak{P}$, then $\mathfrak{P}' \in V$ implies that $d(A_{\mathfrak{P}'}, B_{\mathfrak{Q}'}) = d(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) < d$, hence, $\mathfrak{P}' \in V_d$. If $\mathfrak{P}' \not\supseteq \mathfrak{P}$, $\mathfrak{P}'$ does not contain any of the isolated prime ideal in $Spec(A) - V_d'$, i.e., it is contained in V_d', hence, in V_d. Thus $V'' \cup V_d'$ is an open subset contained in V_d and containing $\mathfrak{P} \notin V_d'$. This contradicts the maximality of V_d', and we can conclude that $V_d' = V_d$, i.e., V_d is open. We get G_d as the complement of V_d in $Spec(A)$.

MAIN THEOREM*.[16] *Let B and $i\colon B \to A$ be as in the above theorem. Then, for each non-negative integer t, there exists an open subset V_t of $Spec(A)$ such that $\mathfrak{P} \in V_t$ if and only if $\mathfrak{P}$ satisfies the condition* $(***)_t$. (See § 1.)

Proof. The statement and the proof of (7), given in the proof of Lemma 3, are independent of the given $\mathfrak{P}$ and the additional assumption: $\operatorname{codh}(A_{\mathfrak{P}}, B_{\mathfrak{Q}}) = 0$. We adopt it here in the same form as it is. Let V_t^* be the open subset of $Spec(A)$ which is the intersection of those open subsets of $Spec(A)$ defined by the ideals $\mathfrak{P}^*$ not satisfying $(***)_t$. It is clear from the definition of $(***)_t$ that if $\mathfrak{P} \supseteq \mathfrak{P}^*$ and if $\mathfrak{P}^*$ does not satisfy $(***)_t$, then $\mathfrak{P}$ does not $(***)_t$. Hence it is necessary for $\mathfrak{P}$ to satisfy $(***)_t$ that $\mathfrak{P} \in V_t^*$. First we shall prove that if $\mathfrak{P} \in V_t^*$, then

(11) *for an $A_{\mathfrak{P}}$-sequence $x_1, \cdots, x_s$ in $B_{\mathfrak{Q}}$ and t' $(\leq t)$ elements*

[16] See the footnote 15. (One can also prove Main Theorem* without assuming the restricted chain condition in B.)

$y_1,\cdots,y_{t'}$ *in* $A_{\mathfrak{P}}$ *such that* $\operatorname{rank}(x_1,\cdots,x_s,y_1,\cdots,y_{t'})A_{\mathfrak{P}}=s+t'$, *the ideal* $(x_1,\cdots,x_s,y_1,\cdots,y_{t'})A_{\mathfrak{P}}$ *is unmixed.*

In proving it, we may assume that $x_1,\cdots,x_s,y_1,\cdots,y_{t'}$ is an $A_{\mathfrak{P}}$-sequence; in fact, if otherwise, there would exist t'' $(0\leqq t''\leqq t'-1)$ such that $x_1,\cdots,x_s,y_1,\cdots,y_{t''}$ is an $A_{\mathfrak{P}}$-sequence but the ideal $(x_1,\cdots,x_s,y_1,\cdots,y_{t''})A_{\mathfrak{P}}$ is not unmixed. Let $\mathfrak{P}''$ be any prime ideal in A such that $\mathfrak{P}''A_{\mathfrak{P}}$ is an associated prime ideal of $(x_1,\cdots,x_s,y_1,\cdots,y_{t'})A_{\mathfrak{P}}$. We want to prove that $\operatorname{rank}(A_{\mathfrak{P}''})=s+t'$. Let us take and fix a $\mathfrak{P}^*$ which has the properties stated in (7) for the above $\mathfrak{P}''$. If $d(A_{\mathfrak{P}^*})$ is proved to be zero, then $d(A_{\mathfrak{P}''})=d(A_{\mathfrak{P}^*})=0$ (by (7)) and therefore $\operatorname{rank}(A_{\mathfrak{P}''})=\operatorname{codh}(A_{\mathfrak{P}''})=s+t'$. (See that $x_1,\cdots,x_s,y_1,\cdots,y_{t'}$ is a maximal $A_{\mathfrak{P}''}$-sequence.) We shall prove that $d(A_{\mathfrak{P}^*})=0$. First of all, the following is clear:

$$(12)\quad \operatorname{codh}(A_{\mathfrak{P}''})=s+t'\leqq \operatorname{codh}(A_{\mathfrak{P}''},B_{\mathfrak{Q}''})+t'.$$

Since $A_{\mathfrak{P}''}/\mathfrak{P}^*A_{\mathfrak{Q}''}$ satisfies the condition (!) over $B_{\mathfrak{Q}''}/\mathfrak{Q}^*B_{\mathfrak{Q}''}$ (see (7)),

$$(13)\quad \operatorname{corank}(\mathfrak{P}^*A_{\mathfrak{P}''})\geqq \operatorname{corank}(\mathfrak{Q}^*B_{\mathfrak{Q}''}).$$

Since $d(A_{\mathfrak{P}''})=d(A_{\mathfrak{P}^*})$ (see (7)), we get

$$\operatorname{codh}(A_{\mathfrak{P}''})-\operatorname{codh}(A_{\mathfrak{P}^*})=\operatorname{rank}(A_{\mathfrak{P}''})-\operatorname{rank}(A_{\mathfrak{P}^*})\geqq \operatorname{corank}(\mathfrak{P}^*A_{\mathfrak{P}''}),$$

hence,

$$\begin{aligned}(14)\quad \operatorname{codh}(A_{\mathfrak{P}^*}) &\leqq \operatorname{codh}(A_{\mathfrak{P}''})-\operatorname{corank}(\mathfrak{P}^*A_{\mathfrak{P}''})\\ &\leqq \operatorname{codh}(A_{\mathfrak{P}''},B_{\mathfrak{Q}''})+t'-\operatorname{corank}(\mathfrak{P}^*A_{\mathfrak{P}''}) &&\text{by (12)}\\ &\leqq \operatorname{codh}(A_{\mathfrak{P}''},B_{\mathfrak{Q}''})+t'-\operatorname{corank}(\mathfrak{Q}^*B_{\mathfrak{Q}''}) &&\text{by (13)}\\ &\leqq \operatorname{codh}(A_{\mathfrak{P}^*},B_{\mathfrak{Q}^*})+t' &&\text{by (2)}.\end{aligned}$$

On the other hand, $\mathfrak{P}\in V_t^*$ and $\mathfrak{P}\supseteq\mathfrak{P}''\supseteq\mathfrak{P}^*$. Therefore, by the definition of V_t^*, $\mathfrak{P}^*$ satisfies the condition $(***)_t$. Hence, if $\operatorname{rank}(A_{\mathfrak{P}^*})>\operatorname{rank}(B_{\mathfrak{Q}^*})+t$, then $\operatorname{codh}(A_{\mathfrak{P}^*})\geqq\operatorname{rank}(B_{\mathfrak{Q}^*})+t+1$, which is impossible by the above result. Therefore $\operatorname{rank}(A_{\mathfrak{P}^*})\leqq\operatorname{rank}(B_{\mathfrak{Q}^*})+t$, and then $(***)_t$ says that $\operatorname{codh}(A_{\mathfrak{P}^*})=\operatorname{rank}(A_{\mathfrak{P}^*})$, i.e., $d(A_{\mathfrak{P}^*})=0$. Thus the assertion (11) on the open subset V_t^* is proved. Let G_1 be the closed subset of *Spec*(A) obtained in Theorem 1, and let $V_1'=\mathit{Spec}(A)-G_1$. As is easily seen, if $\mathfrak{P}$ satisfies $(***)_t$, $\mathfrak{P}\in V_1'$. Setting $V_t=V_1'\cap V_t^*$, we can prove that $\mathfrak{P}\in V_t$ if and only if $\mathfrak{P}$ satisfies the condition $(***)_t$. In fact, the if part has been already proved, and the only-if part follows from the following lemma, in virtue of the above result (11).

LEMMA 4. *Let B' be a local ring and A' a local B'-algebra dominating B'; namely, A' is a local ring with a homomorphism $i\colon B' \to A'$ under which the maximal ideal of B' is the complete inverse image of that of A'. If* $\operatorname{codh}(A', B') = \operatorname{rank}(B')$ (i.e., $d(A', B') = 0$), *then every finite system of elements in B', which generates an ideal in B' of rank = the length of the system, is an A'-sequence.*

Proof. The proof will be an immediate consequence of the results ($''$) of § 1 and $(')_1$ cited in the proof of Lemma 2. Let $x_1, \cdots, x_s$ be any finite system of elements of B' such that $\operatorname{rank}(x_1, \cdots, x_s)B' = s$. Then we can extend it to a system of parameters of B', say $x_1, \cdots, x_s, x_{s+1}, \cdots, x_m$. Let $S = \{x_1, \cdots, x_m\}$. If $x_{1'}, \cdots, x_{m'}$ is a maximal A'-sequence in S, then one of the associated prime ideals of $(x_{1'}, \cdots, x_{m'})A'$ contains S (more rigorously, $i(S)$). Hence it also contains the image of the maximal ideal of B', which implies that $x_{1'}, \cdots, x_{m'}$ is a maximal A'-sequence in B'. However, $m = \operatorname{rank}(B') = \operatorname{codh}(A', B')$, by assumption, and therefore $x_{1'}, \cdots, x_{m'}$ contains all the elements in S. By $(')_1$, $x_1, \cdots, x_m$ is an A'-sequence, hence obviously, $x_1, \cdots, x_s$ is an A'-sequence.

3. Proofs of Proposition and Theorem 2. First, as for the proposition stated in § 1, the existence of B' is not essential and we restate and prove it as follows:

PROPOSITION. *Let A' be a local ring and $\mathfrak{J}'$ an ideal in A'. Let $x_1, \cdots, x_s$ be a system of elements in A', which is an $A'/\mathfrak{J}'$-sequence. For an ideal $\mathfrak{J}'_0$ in A', $\mathfrak{J}'_0\bar{A}' = \mathfrak{J}'_0\bar{A}'$ and $\mathfrak{J}'_0 \subseteq \mathfrak{J}'$ imply $\mathfrak{J}'_0 = \mathfrak{J}'$, where $\bar{A}' = A'/(x_1, \cdots, x_s)A'$.*

Proof. The proof will be by induction on s. First consider the case in which $s = 1$. What is asserted in this case is: *If x_1 is neither a unit of A' nor a zero-divisor in $A'/\mathfrak{J}'$, then $\mathfrak{J}' + x_1A' = \mathfrak{J}'_0 + x_1A'$ and $\mathfrak{J}'_0 \subseteq \mathfrak{J}'$ imply $\mathfrak{J}'_0 = \mathfrak{J}'$.* Since A' is a local ring and x_1 is not a unit of A' we have $\mathfrak{J}'_0 = \bigcap_{n=1}^{\infty} (\mathfrak{J}'_0 + x_1{}^nA')$. Therefore it is sufficient to prove that $\mathfrak{J}' \subseteq \mathfrak{J}'_0 + x_1{}^nA'$ for all positive integers n. This can be proved by induction on n as follows: For $n = 1$, it is included in the assumption. Then, assume that $\mathfrak{J}'_0 + x_1{}^{n-1}A' \supseteq \mathfrak{J}'$ for any $n > 1$. Let y be any element of $\mathfrak{J}'$. Then $y = y_0 + x_1{}^{n-1}z$ with $y_0 \in \mathfrak{J}'_0$ and $z \in A'$. Then $z \in (\mathfrak{J}' : x_1{}^{n-1}A')$. But, since x_1 is not a zero-divisor in $A'/\mathfrak{J}'$, we have $(\mathfrak{J}' : x_1{}^{n-1}A') = \mathfrak{J}'$. Hence, by the assumption in Proposition, $z = y'_0 + x_1z'$ with $y'_0 \in \mathfrak{J}'_0$ and $z' \in A'$, and therefore

$$y = y_0 + x_1^{n-1}z = (y_0 + y'_0) + x_1^n z' \in \mathfrak{J}'_0 + x_1^n A'.$$

Thus the proposition for $s = 1$ is proved. Assume that $s > 1$ and that the proposition is proved for smaller s. Taking $\mathfrak{J}' + x_1A'$, $\mathfrak{J}'_0 + x_1A'$ and $x_2, \cdots, x_s$ in the places of $\mathfrak{J}'$, $\mathfrak{J}'_0$ and $x_1, \cdots, x_s$, we can easily check that the induction assumption is applicable to them and asserts the equality $\mathfrak{J}' + x_1A' = \mathfrak{J}'_0 + x_1A'$. By the above result for $s = 1$, $\mathfrak{J}' = \mathfrak{J}'_0$ follows.

Let $\boldsymbol{V}$ be a normal variety, defined over an algebraically closed field k. Let $\boldsymbol{T}$ be a non-singular variety, defined over k, and $\boldsymbol{Y}$ a $\boldsymbol{V} \times \boldsymbol{T}$-cycle of codimension one. We know that the cycle $\boldsymbol{X}_t = \mathrm{pr}_{\boldsymbol{V}}[\boldsymbol{Y} \cdot \boldsymbol{V} \times \boldsymbol{t}]$ on $\boldsymbol{V}$ is defined for almost all $\boldsymbol{t} \in \boldsymbol{T}$, i.e., for all $\boldsymbol{t}$ in a certain non-empty open subset of $\boldsymbol{T}$. We want to prove that *there exists a non-empty open subset $\boldsymbol{T}'$ of $\boldsymbol{T}$ such that: 1) $\boldsymbol{X}_t$ is defined for every $\boldsymbol{t} \in \boldsymbol{T}'$ and 2) either all or none of the X_t with $\boldsymbol{t} \in \boldsymbol{T}'$ are locally principal (i.e., divisors on $\boldsymbol{V}$ in Cartier's sense).* For this assertion we may assume that $\boldsymbol{T}$ is affine, without any loss of generality. Also, without any loss of generality, we may assume that $\boldsymbol{V}$ is affine, by the fact that a variety admits a finite covering of affine open subsets, and then that the cycle $\boldsymbol{Y}$ is positive, because one can find a principal cycle $\boldsymbol{Y}_0$ of codimension one on the affine variety $\boldsymbol{V} \times \boldsymbol{T}$ such that $\boldsymbol{Y} + \boldsymbol{Y}_0$ is positive.

We can now formulate the problem in the following generality. Let T be a regular integral domain, and W a T-algebra of finite type such that $i\colon T \to W$ is injective. Let $\mathfrak{J}_0$ be an ideal in the rings of quotients V_0 of W with respect to $T^* = T - (0)$. We assume that *all the associated prime ideals of $\mathfrak{J}_0$ have the same rank, say $r > 0$.* Then the intersection $\mathfrak{J} = \mathfrak{J}_0 \cap W$ has the same property. If $\mathfrak{t}$ is a prime ideal in T such that $\mathrm{rank}(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{t}} = \mathrm{rank}(\mathfrak{t}) + r$, then we set $V_{\mathfrak{t}} = W_{\mathfrak{t}}/\mathfrak{t}W_{\mathfrak{t}}$ and $\mathfrak{J}_{\mathfrak{t}} = \mathfrak{J}V_{\mathfrak{t}}$ $(= (\mathfrak{J}, \mathfrak{t})W_{\mathfrak{t}}/\mathfrak{t}W_{\mathfrak{t}})$ and we say that *the specialization $\mathfrak{J}_{\mathfrak{t}}$ of $\mathfrak{J}_0$ is defined.* One can prove that if $\mathfrak{J}_{\mathfrak{t}}$ is defined, then rank $(\mathfrak{J}_{\mathfrak{t}}) \geqq r$. (Note: if $s = \mathrm{rank}(\mathfrak{t})$, $\mathfrak{t}W_{\mathfrak{t}}$ is generated by s elements because T is a regular ring.) If $\mathfrak{J}_{\mathfrak{t}}$ is defined, we also define and denote by $\mathfrak{J}(\mathfrak{t})$ *the reduced specialization of $\mathfrak{J}_0$ at* $\mathfrak{t}$ as follows: $\mathfrak{J}(\mathfrak{t})$ = *the intersection of those primary components of $\mathfrak{J}_{\mathfrak{t}}$ which belong to prime ideals of rank r.*

Remark. Let $\boldsymbol{V}$, $\boldsymbol{T}$ and $\boldsymbol{Y}$ be a normal variety, a non-singular variety and a positive $\boldsymbol{V} \times \boldsymbol{T}$-cycle of codimension one respectively. Let k be a field over which $\boldsymbol{V}$ and $\boldsymbol{T}$ are defined and $\boldsymbol{Y}$ is rational. Let T be an affine ring ring over $\boldsymbol{T}$ over k, and W such of $\boldsymbol{V} \times \boldsymbol{T}$ which contains T. Let $\mathfrak{J}$ be the ideal of $\boldsymbol{Y}$ in W. Let $\boldsymbol{t}$ be a point of $\boldsymbol{T}$ and $\mathfrak{t}$ the ideal of $\boldsymbol{t}$ in T, assuming that $\boldsymbol{t}$ is contained in the open subset of $\boldsymbol{T}$ with which the ring T is associated.

The ring of quotients V_0 of W with respect to $T-(0)$ is an affine ring of $\boldsymbol{V}$, and $\mathfrak{J}V_0$ is the ideal of a *generic* cycle of the family $\{\mathbf{X}_t\}$ with reference to field k. Set $\mathfrak{J}_0 = \mathfrak{J}V_0$. Then we can see that the cycle $\mathbf{X}_t$ is defined if and only if $\mathfrak{J}_0 \cap W_{\mathfrak{t}} = \mathfrak{J}W_{\mathfrak{t}}$ (and $\mathfrak{J}_{\mathfrak{t}}$ is defined) for all W as above. (The second condition in the parenthesis follows from the first, because $\operatorname{rank}(\mathfrak{J}) = 1$.) Furthermore, it is easy to see that there exists an affine ring T of $\boldsymbol{T}$ such that $\mathfrak{J}_0 \cap W = \mathfrak{J}$ for all W containing T, hence, if $\mathfrak{t}$ is a prime ideal in T, then $\mathfrak{J}_0 \cap W_{\mathfrak{t}} = \mathfrak{J}W_{\mathfrak{t}}$ for all W as above. Let $\boldsymbol{t}$ and $\mathfrak{t}$ be as above. Suppose $\mathbf{X}_t$ (hence $\mathfrak{J}_{\mathfrak{t}}$) be defined. Then *the reduced specialization* $\mathfrak{J}(\mathfrak{t})$ *is the ideal of the cycle* $\mathbf{X}_t$ *in the affine ring* $V_{\mathfrak{t}}$ *of* $\boldsymbol{V}$.

In general, an ideal $\mathfrak{J}'$ in a Noetherian ring W' will be said to be *locally principal* if $\operatorname{rank}(\mathfrak{J}') = r > 0$ and $\mathfrak{J}'$ is generated by r elements.

THEOREM 2. *Let T be a regular integral domain, W a T-algebra of finite type such that $i: T \to W$ is injective and $\mathfrak{J}_0$ an ideal in the ring of quotients V_0 of W with respect to $T-(0)$. Let us assume that all the isolated prime ideals in $W_{\mathfrak{P}}$ have the same corank for every prime ideal $\mathfrak{P}$ in W and that all the associated prime ideals of $\mathfrak{J}_0$ have the same rank $r > 0$. Let $\mathfrak{J}_{\mathfrak{t}}$ and $\mathfrak{J}(\mathfrak{t})$ be the symbols which were introduced above and called respectively the specialization and the reduced specialization of $\mathfrak{J}_0$ at $\mathfrak{t} \in Spec(T)$. Then there exists a non-empty open subset U of $Spec(T)$ such that:* i) *$\mathfrak{J}_{\mathfrak{t}}$ (hence $\mathfrak{J}(\mathfrak{t})$) are defined for all $\mathfrak{t} \in U$,* ii) *$\mathfrak{J}_{\mathfrak{t}} = \mathfrak{J}(\mathfrak{t})$ for all $\mathfrak{t} \in U$, and* iii) *either all or none of the $\mathfrak{J}_{\mathfrak{t}}$ with $\mathfrak{t} \in U$ are locally principal.*

Proof. Let $\mathfrak{J} = \mathfrak{J}_0 \cap W$ and $A = W/\mathfrak{J}$. Then the T-algebra A satisfies the conditions (on A with $B = T$) in the corollary to Main Theorem for $t = 0$; namely, beside basic assumptions on T and A, the ring of quotients A_0 of A with respect to $T-(0)$ satisfies the condition $(-)_0$. By the corollary, there exists a non-empty open subset U_0 of $Spec(T)$ such that, every $\mathfrak{t} \in U_0$, $\mathfrak{t}A_{\mathfrak{t}} \neq A_{\mathfrak{t}}$ and a regular system of parameters of $T_{\mathfrak{t}}$ is an $A_{\mathfrak{p}}$-sequence and generates an unmixed ideal in $A_{\mathfrak{p}}$ if $\mathfrak{p} \in Spec(A)$ and $\mathfrak{p} \cap T = \mathfrak{t}$. We shall prove that all $\mathfrak{t} \in U_0$ possess the properties i) and ii). Let $\mathfrak{t} \in U_0$ and take a regular system of parameters, $x_1, \cdots, x_s$, of $T_{\mathfrak{t}}$ ($s = \operatorname{rank}(\mathfrak{t})$). Then the above result implies that: (a) $(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{t}} \neq W_{\mathfrak{t}}$ and (b) $(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{P}}$ $(= (\mathfrak{J}, x_1, \cdots, x_s)W_{\mathfrak{P}})$ is unmixed and $(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{P}}/\mathfrak{J}W_{\mathfrak{P}}$ is of rank s for every $\mathfrak{P} \in Spec(W)$ such that $\mathfrak{P} \supseteq \mathfrak{J}$ and $\mathfrak{P} \cap T = \mathfrak{t}$. Since W is a homomorphic image of a regular ring (that is, a polynomial ring over T), it satisfies the *restricted* chain condition on the prime ideals. (See the definition preceding Lemma 3.) By the assumption that all the isolated prime ideals of zero in $W_{\mathfrak{P}}$ have the same corank, $W_{\mathfrak{P}}$ satisfies *the chain condition on the*

prime ideals, i. e., all the maximal chain of prime ideals in $W_{\mathfrak{P}}$ have the same length. In (b), if $\mathfrak{P}$ is an associated prime ideal of $(\mathfrak{J}, \mathfrak{t})$ then $\operatorname{rank}(W_{\mathfrak{P}}/\mathfrak{J}W_{\mathfrak{P}}) = s$ and, by the above chain condition,

$$\operatorname{rank}(W_{\mathfrak{P}}) = s + \operatorname{rank}(\mathfrak{J}W_{\mathfrak{P}}) - \operatorname{rank}(\mathfrak{t}) + r.$$

(Recall the assumption that all the associated prime ideals of $\mathfrak{J}_0$, hence of $\mathfrak{J}$, have the same rank r.) In other words, all the associated prime ideals of $(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{t}}$ have the same rank $= \operatorname{rank}(\mathfrak{t}) + r$. The statements i) and ii) follow immediately. We now want to find a non-empty open subset U_0 of U for which iii) holds. We shall consider the following two cases separately: one in which $\mathfrak{J}_0$ is locally principal and the other in which $\mathfrak{J}_0$ is not. Suppose $\mathfrak{J}_0$ be locally principal. First of all, we can easily see that there exists an open subset V of $Spec(W)$ such that $\mathfrak{P} \in V$ if and only if either $\mathfrak{J}W_{\mathfrak{P}} = W_{\mathfrak{P}}$ or $\mathfrak{J}W_{\mathfrak{P}}$ is principal, i. e., generated by r elements. By Chevalley's theorem, the image of $Spec(W) - V$ in $Spec(T)$ is a *constructible* subset of $Spec(T)$. Since $\mathfrak{J}_0 = \mathfrak{J}V_0$ is locally principal, this constructible subset of $Spec(T)$ does not contain any open non-empty subset of $Spec(T)$. Therefore there exists a non-empty open subset of $Spec(T)$, say U', which does not intersect with the image of $Spec(W) - V$. It is clear that the non-empty open subset $U_0 \cap U' = U$ is such that, in addition to i) and ii), all the $\mathfrak{J}_{\mathfrak{t}}$ with $\mathfrak{t} \in U$ are locally principal. Let us now consider the other case in which $\mathfrak{J}_0$ is not locally principal. Then there exist at least one prime ideal $\mathfrak{P}$ in W such that $\mathfrak{P} \cap T = (0)$, i. e., $\mathfrak{P}V_0 \in Spec(V_0)$ and that $\mathfrak{J}W_{\mathfrak{P}}$ is not principal, i. e., not generated by r elements, and $\mathfrak{J}W_{\mathfrak{P}} \neq W_{\mathfrak{P}}$. Since $\mathfrak{P} \cap T = (0)$, the canonical homomorphism of T into the T-algebra $W/\mathfrak{P}$ is injective and therefore, again in virtue of the theorem of Chevalley, there can be found a non-empty open subset U'' of $Spec(T)$ which is contained in the image of $Spec(W/\mathfrak{P})$ in $Spec(T)$. Obviously, this open subset U'' has the property that for every $\mathfrak{t}'' \in U''$ there exists at least one $\mathfrak{P}'' \in Spec(W)$ such that $\mathfrak{P}'' \cap T = \mathfrak{t}''$, $\mathfrak{J}W_{\mathfrak{P}''} \neq W_{\mathfrak{P}''}$ and $\mathfrak{J}W_{\mathfrak{P}''}$ is not principal; in fact, we have $\mathfrak{P}'' \in Spec(W)$ such that $\mathfrak{P}'' \cap T = \mathfrak{t}''$ and $\mathfrak{P}'' \supseteq \mathfrak{P} \supseteq \mathfrak{J}$. Let $U = U_0 \cap U''$, which is a non-empty open subset of $Spec(T)$. If $\mathfrak{t}'' \in U$, such $\mathfrak{P}'' \in Spec(W)$ as above possesses moreover the property that a regular system of parameters of $T_{\mathfrak{t}''}$ is an $A_{\mathfrak{p}}$-sequence, where $A = W/\mathfrak{J}$ and $\mathfrak{p} = \mathfrak{P}''/\mathfrak{J}$, hence, a $W_{\mathfrak{P}''}/\mathfrak{J}W_{\mathfrak{P}''}$-sequence. (It follows from the definition of U_0.) By Proposition, where $A' = W_{\mathfrak{P}''}$ and $\mathfrak{J}' = \mathfrak{J}W_{\mathfrak{P}''}$, if $(\mathfrak{J}, \mathfrak{t})W_{\mathfrak{P}''}/\mathfrak{t}W_{\mathfrak{P}''}$ is generated by r elements then $\mathfrak{J}W_{\mathfrak{P}''}$ is also generated by r elements. But this should not be true. Hence $\mathfrak{J}_{\mathfrak{t}''}$ $(= (\mathfrak{J}, \mathfrak{t})W_{\mathfrak{t}}/\mathfrak{t}W_{\mathfrak{t}})$ is not locally principal. (Note: $W_{\mathfrak{P}''}/\mathfrak{t}W_{\mathfrak{P}''}$ is identified

with a local ring of W_t/tW_t.) Again the open subse U of $Spec(T)$ has the properties required in the theorem.

REFERENCES.

[1] J. Dieudonné and A. Grothendieck, *Élements de géometrie algébrique*, forthcoming in mathematical publications of l'Institut des Hautes Edues Scientifiques.

[2] W. Krull, "Primidealketten in allgemeinen Ringbereiche," *Sitzungsberichte der Heidelberger Akademie der Wissenschaften*, vol. 7 (1928), pp. ———.

[3] ———, "Parameterspecializierung in Polynomringen," *Archiv der Mathematik*, Band 1 (1948), pp. 56-64.

[4] A. Seidenberg, "The hyperplane sections of normal varieties," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 357-386.

[5] M. Auslander and D. A. Buchsbaum, "Codimension and multiplicity," *Annals of Mathematics*, vol. 68 (1958), pp. 625-657.

[6] ———, "Homological dimension in local rings," *Transactions of the American Mathematical Society*, vol. 85 (1957), pp. 390-405.

[7] A. Weil, *Foundations of algebraic Geometry*, American Mathematical Society Colloquium Publications, 1946.

THE WEDDERBURN DECOMPOSITION OF COMMUTATIVE BANACH ALGEBRAS.*

By WILLIAM G. BADÉ and PHILIP C. CURTIS, JR.[1]

I. Let $\mathfrak{A}$ be a Banach algebra with radical $\mathfrak{R}$. If the dimension of $\mathfrak{A}$ is finite, the Wedderburn principal theorem states that there exists a subalgebra $\mathfrak{B}$ of $\mathfrak{A}$ such that $\mathfrak{A} = \mathfrak{B} + \mathfrak{R}$ and $\mathfrak{B} \cap \mathfrak{R} = \{0\}$. Moreover, if $\mathfrak{A}$ is commutative, $\mathfrak{B}$ is necessarily unique. It is the purpose of this paper to investigate under what conditions this theorem will hold for infinite dimensional commutative Banach algebras. In 1954 Zelinsky [17] gave an example which showed that the Wedderburn principal theorem did not hold in general for infinite dimensional algebras. No topological assumptions were made in this construction, and in particular the algebra is not a Banach algebra. In 1955 Feldman [6] constructed an example of a commutative Banach algebra with one dimensional radical but with no closed subalgebra $\mathfrak{B}$ such that $\mathfrak{A} = \mathfrak{B} + \mathfrak{R}$ and $\mathfrak{B} \cap \mathfrak{R} = \{0\}$. Another example with similar properties is given by Glaeser [8]. Various condition under which the theorem does hold for closed subalgebras $\mathfrak{B}$ are discussed by both Feldman and Glaeser.

In this paper we shall also be concerned with the validity of the Wedderburn theorem for non-closed subalgebras $\mathfrak{B}$, so the following terminology will be helpful. We shall call a Banach algebra *decomposable* if there exists a subalgebra $\mathfrak{B}$ of $\mathfrak{A}$ with the property that $\mathfrak{A} = \mathfrak{B} + \mathfrak{R}$ and $\mathfrak{B} \cap \mathfrak{R} = \{0\}$ ($\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$). The decomposition is unique if there is only one such subalgebra. If there exists a closed subalgebra $\mathfrak{B}$ such that $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$, then $\mathfrak{A}$ will be said to be *strongly decomposable*. A strong decomposition is unique if $\mathfrak{B}$ is unique among closed subalgebras of $\mathfrak{A}$.

Unfortunately, nothing but the obvious relations between these notions holds in general for commutative Banach algebras. In [4]† the present authors showed that Feldman's example was decomposable even though not strongly

* Received February 23, 1960.

[1] This work was supported by the Office of Naval Research under contract Nonr 233 (59), the Air Force Office of Scientific Research under contract SAR/AF-49(636)-153 and by the National Sciences Foundation.

† Erratum to [4]: On page 604, line 11, read $\overline{y(F)}$ for $y(F)$.

so. A simple example discussed in § V of this paper shows that a commutative Banach algebra may have a non-unique decomposition but a unique strong decomposition. A more illuminating example, also discussed in § V, shows that a commutative algebra need not be decomposable at all. In fact, this example is isomorphic modulo the radical to the ring of convergent sequences.

On the positive side of the ledger, however, some fairly general criteria can be given under which a commutative algebra is decomposable. Many of the results are valid for mildly non-commutative algebras as well. As might be expected from the finite dimensional situation, most of the results depend on an abundance of idempotents. For convenience we shall always assume that the algebra possesses a unit. The basic criteria for decomposability, proved in § II, is the following: If A has totally disconnected maximal ideal space $\Phi_{\mathfrak{A}}$, then $\mathfrak{A}$ is decomposable and $\mathfrak{A}/\mathfrak{R} \simeq C(\Phi_{\mathfrak{A}})$ [2] if and only if the idempotents of $\mathfrak{A}$ form a bounded set. If the latter condition holds, then $\mathfrak{A}$ is strongly decomposable and the strong decomposition is unique. Actually, if $\mathfrak{A}/\mathfrak{R} \simeq C(\Phi_{\mathfrak{A}})$, we know of no examples for which the decomposition fails to be unique. The existence of a non closed complement to the radical is equivalent to the existence of a discontinuous isomorphism of $C(\Phi_{\mathfrak{A}})$ into $\mathfrak{A}$, and no such examples are known (cf. [4]).

In § III it is proved that the set of idempotents of A will automatically be bounded if the idempotents satisfy a weak interpolation axiom. This axiom will be satisfied if and only if $\Phi_{\mathfrak{A}}$ is a totally disconnected F space in the sense of Gillman and Henriksen. Thus, for such spaces $\Phi_{\mathfrak{A}}$, which include extremally disconnected compact spaces, we can give a virtually complete Wedderburn structure for the algebra $\mathfrak{A}$. More precisely, for a commutative Banach algebra $\mathfrak{A}$, if $\Phi_{\mathfrak{A}}$ is a totally disconnected F space, then $\mathfrak{A}/\mathfrak{R} \simeq C(\Phi_{\mathfrak{A}})$, $\mathfrak{A}$ is strongly decomposable and the strong decomposition is unique. An obvious but interesting corollary is that if Ω is a totally disconnected F space, then $C(\Omega)$ is the only commutative semi-simple Banach algebra with Ω as maximal ideal space, e.g. l_∞ is the only semi-simple Banach algebra having the Čech compactification of the integers as maximal ideal space.

In § IV we investigate conditions on the radical which will insure decomposability of $\mathfrak{A}$ under the assumption that $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$. For an arbitrary compact space Ω, $\mathfrak{A}$ will be strongly decomposable if the dimension of $\mathfrak{R}$ is finite. If Ω is totally disconnected, this will be the case if $\mathfrak{R}$ is nilpotent. In both cases the decomposition is unique. Neither the assumption

[2] The symbol $\simeq$ will always mean an algebraic isomorphism. No topological conditions are assumed.

that $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ or $\mathfrak{R}^n = \{0\}$ alone is sufficient to guarantee decomposability, however. Examples illustrating these facts are discussed in § V.

II. *Let* $\mathfrak{A}$ be a Banch algebra which we will assume possesses a unit denoted by e. This is only a convenience, for if $\mathfrak{A}$ has no unit and $\mathfrak{A}'$ is formed in the usual way by adjoining one to $\mathfrak{A}$, then $\mathfrak{A}$ is (strongly) decomposable iff $\mathfrak{A}'$ is (strongly) decomposable. Further, a decomposition or a strong decomposition is unique in $\mathfrak{A}$ iff it is unique in $\mathfrak{A}'$. We leave the verification of these statements to the reader.

We will denote by $\mathfrak{R}$ the Jacobson radical[3] of $\mathfrak{A}$. $\mathfrak{A}/\mathfrak{R}$ is a semisimple Banach algebra under the usual coset norm $\|\bar{x}\| = \inf_{r \in \mathfrak{R}} \|x + r\|$, where $\bar{x} = x + \mathfrak{R}$. Unless otherwise stated this will always be the norm taken in $\mathfrak{A}/\mathfrak{R}$. If $x \in \mathfrak{R}$, then $\lim_{n\to\infty} \| x^n \|^{1/n} = 0$; and if A is commutative, this property characterizes the elements of $\mathfrak{R}$. We will denote by ν the natural homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{R}$. ν is continuous, and if there exists a subalgebra $\mathfrak{B}$ such that $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$, then ν maps $\mathfrak{B}$ isomorphically onto $\mathfrak{A}/\mathfrak{R}$.

Let ν' be the inverse of the restriction of ν to $\mathfrak{B}$. ν' is an isomorphism of $\mathfrak{A}/\mathfrak{R}$ onto $\mathfrak{B}$, and it is easily verified that ν' is continuous iff $\mathfrak{B}$ is closed. If $\mathfrak{B}$ is closed, then $\mathfrak{A}$ is the topological direct sum of $\mathfrak{B}$ and $\mathfrak{R}$ by the well known theorem of Kober [12], and $\mathfrak{B}$ is homeomorphic with $\mathfrak{A}/\mathfrak{R}$.

Next we collect a few elementary facts from algebra. They are valid in a more general setting, but the following formulation is sufficient for our needs. Although we are concerned primarily with commutative algebras, many of the results of this section will be stated for mildly non-commutative algebras to facilitate their use in § IV.

LEMMA 2.1. *Let $P_{\mathfrak{A}}$ denote the set of idempotents of a Banach algebra $\mathfrak{A}$. Then* (1) $P_{\mathfrak{A}} \cap R = \{0\}$. (2) *If $x \equiv x^2 (\mathfrak{R})$, then there exists $p \in P_{\mathfrak{A}}$, satisfying $x \equiv p (\mathfrak{R})$, and $yp = py$ for any $y \in \mathfrak{A}$ such that $yx = xy$.* (3) *If $p, q \in P_{\mathfrak{A}}$, and $pq = qp$, then $p \equiv q (\mathfrak{R})$ implies $p = q$.*

Proof. (1) is immediate since $\mathfrak{R}$ consists of quasi regular elements, and 0 is the only quasi regular idempotent. For the proof of (2) we refer the reader to [15, Theorem 2.3.9]. Lastly, if $p, q \in P_{\mathfrak{A}}$ and $pq = qp$, then $p - q \in \mathfrak{R}$ implies $p(p-q), -q(p-q) \in P_{\mathfrak{A}} \cap \mathfrak{R}$. Hence $p(p-q) = 0 = -q(p-q)$ or $p = q$.

[3] For a full discussion of the properties of the radical the reader is referred to [15, chapter 2].

LEMMA 2.2. *If $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$ for some subalgebra $\mathfrak{B}$ of $\mathfrak{A}$, and if $pq = qp$ for each $p, q \in P_{\mathfrak{A}}$, then $P_{\mathfrak{A}} \subset \mathfrak{B}$.*

Proof. For each $x \in \mathfrak{A}$ there is a unique representation $x = y_x + k$, $y_x \in \mathfrak{B}$, $k \in \mathfrak{R}$. The mapping $\tau: x \to y_x$ is a homomorphism of $\mathfrak{A}$ onto $\mathfrak{B}$. Therefore, if $p \in P_{\mathfrak{A}}$, then $q = \tau(p) \in P_{\mathfrak{A}} \cap \mathfrak{B}$ and $p \equiv q(\mathfrak{R})$. By Lemma 2.1 (3) $p = q$, and thus $P_{\mathfrak{A}} \subset \mathfrak{B}$.

LEMMA 2.3. *If $pr = rp$ for all $p \in P_{\mathfrak{A}}$, $r \in \mathfrak{R}$, and if $\mathfrak{A}/\mathfrak{R}$ is commutative, then $pq = qp$, $p, q \in P_{\mathfrak{A}}$.*

Proof. Let $p, q \in P_{\mathfrak{A}}$. We note first that if $pq \in \mathfrak{R}$, then $pq = qp = 0$. For by hypothesis $pq \in \mathfrak{R}$ implies $pqp = pq$, and hence $(pq)^2 = pq = 0$ by Lemma 2.1(1). Since $\mathfrak{A}/\mathfrak{R}$ is commutative, $pq \in \mathfrak{R}$ implies $qp \in \mathfrak{R}$, and the assertion follows. In general, by Lemma 2.1(2) there is a $q_0 \in P_{\mathfrak{A}}$ such that $pq \equiv q_0 \equiv qp(\mathfrak{R})$. Clearly $(1-p)q_0$ and $(1-q)q_0 \in \mathfrak{R}$. Thus, by what we have just shown, $q_0 = pq_0 = q_0p = qq_0 = q_0q$. Therefore $p - q_0, q - q_0 \in P_{\mathfrak{A}}$, and since $(p-q_0)(q-q_0) \in \mathfrak{R}$, an application of the same arguments proves that $pq = q_0 = qp$.

Before proving the main result of this section we note the following fact which is due to Rickart [14, Theorem 1]. If ν is any isomorphism of $C(\Omega)$ [4] into a commutative Banach algebra $\mathfrak{A}$, then each ω may be extended to a multiplicative linear functional on $\mathfrak{A}$. This yields two corollaries of use to us. One, a result of Kaplansky [11, Theorem 6.2], is that $\| \nu(x) \| \geqq \sup_{\omega \in \Omega} | x(\omega) |$. Hence ν is a homeomorphism iff the range of ν is closed. Secondly, if $\mathfrak{A} = C(\Omega')$ and ν is continuous, then ν maps onto $\mathfrak{A}$ if $\nu(C(\Omega))$ separates the points of Ω'. This follows from the Stone-Weierstrass theorem, since under these hypotheses $\nu(C(\Omega))$ must be self adjoint in $C(\Omega')$.

The next result is our main criterion for decomposability.

THEOREM 2.4. *Let $\mathfrak{A}$ be a Banach algebra with radical $\mathfrak{R}$, and let $\mathfrak{B} = \mathfrak{A}/\mathfrak{R}$. Assume that $\mathfrak{B}$ is commutative with totally disconnected maximal ideal space $\Phi_{\mathfrak{B}}$ and assume that $pq = qp$ for $p, q \in P_{\mathfrak{A}}$. Then $P_{\mathfrak{A}}$ is a bounded set iff $\mathfrak{A}$ is decomposable and $\mathfrak{B} \simeq C(\Phi_{\mathfrak{B}})$. If $P_{\mathfrak{A}}$ is a bounded set, then $\mathfrak{A}$ is strongly decomposable and the strong decomposition is unique. If in addition $\mathfrak{R}$ is nilpotent, the decomposition is necessarily strong and unique.*

Proof. Suppose $\mathfrak{A}$ is decomposable and $\mathfrak{B} \simeq C(\Phi_{\mathfrak{B}})$. Let $\mathfrak{B}'$ be a subalgebra of $\mathfrak{A}$ such that $\mathfrak{A} = \mathfrak{B}' \oplus \mathfrak{R}$, and let P_1 denote the set of idempotents of $C(\Phi_{\mathfrak{B}})$. Since $\mathfrak{B}' \simeq \mathfrak{B} \simeq C(\Phi_{\mathfrak{B}})$, there exist an isomorphism μ of $C(\Phi_{\mathfrak{B}})$

[4] The symbols Ω, Ω' will always denote compact Hausdorff spaces.

onto $\mathfrak{B}'$. By Lemma 2.2 $P_{\mathfrak{A}} \subset \mathfrak{B}'$, hence $\mu(P_1) = P_{\mathfrak{A}}$. By [4, Theorem 2.3] $P_{\mathfrak{A}}$ is a bounded set in $\mathfrak{A}$.

Conversely, suppose $P_{\mathfrak{A}}$ is bounded. Let $\mathfrak{B}'$ be the closed commutative subalgebra of $\mathfrak{A}$ generated by $P_{\mathfrak{A}}$. By [5, Theorem 17] there exists a bicontinuous isomorphism τ of $\mathfrak{B}'$ onto $C(\Phi_{\mathfrak{B}'})$. It is an obvious corollary of the result of Rickart mentioned earlier [14, Theorem 1] that for $x \in \mathfrak{B}'$, $\sup_{\phi \in \Phi_{\mathfrak{B}'}} |\tau(x)(\phi)| = r_{\mathfrak{B}'}(x)$.[5] From this it follows that $\mathfrak{B}' \cap \mathfrak{R} = \{0\}$, since $x \in \mathfrak{B}' \cap \mathfrak{R}$ implies $\lim_{n \to \infty} \| x^n \|^{1/n} = r_{\mathfrak{B}'}(x) = 0$.

Let ν be the natural homomorphism of $\mathfrak{A}$ onto $\mathfrak{B} = \mathfrak{A}/\mathfrak{R}$, and let ρ be the Gelfand isomorphism of $\mathfrak{B}$ into $C(\Phi_{\mathfrak{B}})$. If $\mu = \rho\nu$, μ maps $\mathfrak{A}$ into $C(\Phi_{\mathfrak{B}})$ and is one to one on $\mathfrak{B}'$. We assert μ maps $\mathfrak{B}'$ onto $C(\Phi_{\mathfrak{B}})$. By the well known result of Silov [16],[6] $\rho(P_{\mathfrak{B}}) = P_1$ and by Lemma 2.1(2), $\nu(P_{\mathfrak{A}}) = P_{\mathfrak{B}}$. Therefore, $\mu(P_{\mathfrak{A}}) = P_1$ and $\mu(\mathfrak{B}')$ contains the linear span $\mathfrak{C}$ of P_1. Since $\Phi_{\mathfrak{B}}$ is assumed to be totally disconnected, $\mathfrak{C}$ (and therefore $\mu(\mathfrak{B}')$) is dense in $C(\Phi_{\mathfrak{B}})$. Therefore, by [14, Theorem 1], $\mu\tau^{-1}$ maps $C(\Phi_{\mathfrak{B}'})$ onto $C(\Phi_{\mathfrak{B}})$. Consequently, $\mu(\mathfrak{B}') = C(\Phi_{\mathfrak{B}}) = \rho(\mathfrak{B})$, and $A = \mathfrak{B}' \oplus \mathfrak{R}$.

Finally, suppose $\mathfrak{A} = \mathfrak{B}'' \oplus \mathfrak{R}$ for some subalgebra $\mathfrak{B}''$. There exists an isomorphism λ of $C(\Phi_{\mathfrak{B}})$ onto $\mathfrak{B}''$, and since by Lemma 2.2 $\mathfrak{B} \cap \mathfrak{B}'' \supset P_{\mathfrak{A}}$, it follows that $\lambda(P_1) = P_{\mathfrak{A}}$. If $\mathfrak{B}''$ is closed, λ is a homeomorphism and consequently $\mathfrak{B}' = \mathfrak{B}''$. If $\mathfrak{R}$ is nilpotent[7], then by [4, Theorem 4.5] λ must be continuous, and the theorem follows.

We remark that no examples are known for which the decomposition in the above theorem fails to be unique. As is shown in [4] the problem of constructing such an example is equivalent to the problem of constructing a discontinuous isomorhpism of $C(\Omega)$ for a compact space Ω.

III. In this section we shall suppose that $\mathfrak{A}$ is a Banach algebra in which the elements of $P_{\mathfrak{A}}$ commute. Under this hypothesis it then follows that $P_{\mathfrak{A}}$ will automatically be bounded if it satisfies a simple interpolation axiom. This is an extension of an earlier result of Badé [3, Theorem 2.2] to the effect that a σ-complete Boolean algebra of projections in a Banach space is always bounded. One may strip away the Banach space framework from

[5] For a commutative Banach algebra A, and an element $x \in A$,

$$r_A(x) = \sup_{\phi \in \Phi_A} |x(\phi)|.$$

[6] A more accessible reference is [15, Theorem 3.6.3].

[7] Added in proof: The hypothesis of [4, Theorem 4.5] is that $\mathfrak{R}$ is a nil ideal. This involves no more generality than the assumption of nilpotency, for a simple category argument shows that if a commutative Banach algebra is a nil algebra, then it must be nilpotent.

this theorem and obtain a result about general Boolean rings. This we do below except that we weaken σ-completeness to the following interpolation axiom.

Definition. Let A be a Boolean ring. A is said to satisfy property (I) if for any pair of sequences $\{a_n\}, \{b_n\} \subset A$, $n = 1, 2, \cdots$, for which $a_n a_m = b_n b_m = 0$, $n \neq m$, and $a_n b_m = 0$, for all n, m, there exists an element $a \in A$ such that $a a_n = a_n$ and $a b_n = 0$, $n = 1, 2, \cdots$. The following two results give the facts that we need. We do not assume that the Boolean ring possesses a unit, although in the applications this will be the case.

LEMMA 3.1. *Let A be a Boolean ring, ρ a function on A to the non-negative reals with the following property.*

$$\text{(1)} \qquad \text{If } x \geqq y, \text{ then } \rho(x-y) \geqq |\rho(x) - \rho(y)|.$$

If ρ is unbounded on A, then there exists a sequence $\{x_n\} \subset A$ such that $x_n x_m = 0$, $n \neq m$, and $\lim_{n\to\infty} \rho(x_n) = \infty$.

Proof. It follows easily from (1) that if $xy = 0$, $\rho(x+y) \leqq \rho(x) + \rho(y)$. An element $x \in A$ is said to have property (α) if $\sup_{y \leqq x} \rho(y) = \infty$. We prove the theorem first under the assumption that there exist an $x \in A$ with (α). We will construct a sequence $\{y_n\} \subset A$, $n = 1, 2, \cdots$, $y_n \geqq y_{n+1}$, satisfying $\rho(y_n) \geqq n + \rho(y_{n-1})$. If $x_n = y_n - y_{n+1}$, then $\{x_n\}$ satisfies the conclusion of the lemma. Note that if $y \leqq x$ and x has (α), then either y or $(x-y)$ has (α). Let $y_1 = x$ and choose $z_1 \leqq y_1$ for which $\rho(z_1) \geqq 2 + 2\rho(y_1)$. Let y_2 be the member of the pair $(z_1, y_1 - z_1)$ with (α). Then $\rho(y_2) \geqq 2 + \rho(y_1)$. Now choose $z_2 \leqq y_2$ for which $\rho(z_2) \geqq 3 + 2\rho(y_2)$. If y_3 is the member of the pair $(z_2, y_2 - z_2)$ with (α), then $\rho(y_3) \geqq 3 + \rho(y_2)$. Proceeding inductively we construct the desired sequence $\{y_n\}$.

Now assume that for each $x \in A$, $\sup_{y \leqq x} \rho(y) < \infty$. We construct the sequence $\{x_n\}$ directly. Choose x_1 with $\rho(x_1) \geqq 1$ and let $M_1 = \sup_{y \leqq x_1} \rho(y)$. Then there exists $z_1 \in A$ with $\rho(z_1) \geqq 2 + M_1$. If $x_2 = z_1 - x_1 z_1$, $x_1 x_2 = 0$, and $\rho(x_2) = \rho(z_1 - x_1 z_1) \geqq \rho(z_1) - \rho(x_1 z_1) \geqq 2$. Let $M_2 = \sup_{y \leqq x_1 + x_2} \rho(y)$. Choose z_2 for which $\rho(z_2) \geqq 3 + M_2$. If $x_3 = z_2 - (x_1 + x_2) z_2$, $x_3 x_2 = x_3 x_1 = 0$ and $\rho(x_3) \geqq 3$. The construction proceeds inductively.

The proof of the next result is very similar to that of [4, Theorem 2.1]. Suggestions of Y. Katznelson have greatly simplified the original argument.

THEOREM 3.2. *Let A be a Boolean ring, ρ a function from A to the non negative reals satisfying*

1) $x \geqq y \Rightarrow \rho(x-y) \geqq |\rho(x)-\rho(y)|$,

2) $\rho(xy) \leqq \rho(x)\rho(y)$.

If A satises property (I), *then* $\sup_{y \in A} \rho(y) < \infty$.

Proof. Suppose $\sup_{y \in A} \rho(y) = \infty$. Then by Lemma 3.1 there exists $\{x_n\} \subset A$, $n = 1, 2, \cdots$, $x_n x_m = 0$, $n \neq m$, for which $\lim_{n \to \infty} \rho(x_n) = \infty$. Choose distinct elements y_{jk}, $j, k = 1, 2, \cdots$, from the sequence $\{x_n\}$ such that

$$\rho(y_{jk}) > 2^{j+k}.$$

Since A satisfies (I), by induction we may define an orthogonal sequence $\{y_j\} \subset A$ such that $y_j y_{jk} = y_{jk}$, $y_i y_{jk} = 0$, $i \neq j$. For each integer j select an integer k_j large enough so that $2^{k_j} > \rho(y_j)$. The sequences $\{y_{jk_j}\}$, $\{y_j - y_{jk_j}\}$ satisfy the hypothesis of property (I). Therefore there exists $y \in A$ satisfying $yy_{jk_j} = y_{jk_j}$, $y(y_j - y_{jk_j}) = 0$. Therefore $yy_j = yy_{jk_j} = y_{jk_j}$. Applying (2) we have

$$2^{j+k_j} < \rho(y_{jk_j}) \leqq \rho(y)\rho(y_j) < \rho(y)2^{k_j}.$$

Therefore $\rho(y) > 2^j$ for each integer j, which is impossible.

The application of these results to Banach algebras with commuting idempotents is straight forward. For $p, q \in P_{\mathfrak{A}}$ we define $p \oplus q = p + q - 2pq$. Then $P_{\mathfrak{A}}$ is a Boolean ring with respect to this 'symmetric difference' addition and ordinary multiplication. Since for orthogonal idempotents $p \oplus q = p + q$, the norm satisfies condition (1), and we have the following result.

COROLLARY 3.3. *Let $\mathfrak{A}$ be a Banach algebra such that the elements of $P_{\mathfrak{A}}$ commute. Then if $P_{\mathfrak{A}}$ satisfies property* (I), *$P_{\mathfrak{A}}$ is a bounded set.*

If now $\mathfrak{A}$ is commutative and $\Phi_{\mathfrak{A}}$ is totally disconnected, then property (I) for $P_{\mathfrak{A}}$ can be characterized topologically in $\Phi_{\mathfrak{A}}$. In fact, in light of Lemma 2.1 and the theorem of Silov, $P_{\mathfrak{A}}$ satisfies (I) if and only if $P_{C(\Phi_{\mathfrak{A}})}$ satisfies (I). However, it is easily verified [10, p. 1619] that the latter condition holds if and only if disjoint open F_σ sets in $\Phi_{\mathfrak{A}}$ have disjoint closures. Such spaces are called F spaces by Gillman and Henriksen and are extensively studied in [7]. Examples include the Stone spaces of σ-complete Boolean algebras, the Čech compactification of the integers $\beta(N)$, and also $\beta(N) - N$. The latter is not the Stone space of a σ-complete Boolean algebra as is remarked in [10].

Combining these remarks with Theorem 2.4 and Corollary 3.3 we have

THEOREM 3.4. *If $\mathfrak{A}$ is a commutative Banach algebra and $\Phi_{\mathfrak{A}}$ is a totally*

disconnected F space, then $\mathfrak{A}$ *is strongly decomposable and* $\mathfrak{A}/\mathfrak{R} \simeq C(\Phi_{\mathfrak{A}})$. *The strong decomposition is in addition unique.*

COROLLARY 3.5. *If* Ω *is a totally disconnected compact F space, then* $C(\Omega)$ *is the only semi-simple commutative Banach algebra with* Ω *as maximal ideal space.*

It is an open question whether the class of totally disconnected compact F spaces is characterized by the property of Corollary 3.5.

IV. In this section we shall consider commutative algebras $\mathfrak{A}$ such that $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ for a compact space Ω, and give two conditions on the radical which will imply decomposability of $\mathfrak{A}$ for appropriate spaces Ω.

We note first that if σ is an isomorphism of $\mathfrak{A}/\mathfrak{R}$ onto $C(\Omega)$ then Ω is homeomorphic with $\Phi_{\mathfrak{A}}$. This follows easily since σ^{-1} followed by the Gelfand isomorphism of $\mathfrak{A}/\mathfrak{R}$ is a continuous isomorphism μ of $C(\Omega)$ onto a separating subalgebra of $C(\Phi_{\mathfrak{A}})$. By the result of Rickart [14, Theorem 1] discussed in § II, μ maps onto $C(\Phi_{\mathfrak{A}})$, and Ω is therefore homeomorphic with $\Phi_{\mathfrak{A}}$ by the classical result of Stone.

It is shown in § V that $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ by itself is not enough to guarantee the decomposability of $\mathfrak{A}$. However, if $\mathfrak{A}$ is decomposable the following general result holds.

THEOREM 4.1. *Let* $\mathfrak{A}$ *be a commutative Banach algebra such that* $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$. *If* $\mathfrak{A}$ *is decomposable, it is strongly decomposable, and the strong decomposition is unique.*

Proof. Suppose $\mathfrak{A} = \mathfrak{B}' \oplus \mathfrak{R}$ and let σ be the induced isomorphism of $C(\Omega)$ onto $\mathfrak{B}'$. By [4, Theorem 2.3] there exists a continuous isomorphism μ of $C(\Omega)$ into $\mathfrak{A}$ such that $\sigma - \mu$ maps $C(\Omega)$ into $\mathrm{Rad}(\overline{\mathfrak{B}'})$. If $\mathfrak{B} = \mu(C(\Omega))$, then $\mathfrak{B}$ is closed and $\mathfrak{B} \cap \mathfrak{R} = \{0\}$. Now $\mathrm{Rad}(\overline{\mathfrak{B}'}) = \overline{\mathfrak{B}'} \cap \mathfrak{R}$. It follows easily that $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$, so $\mathfrak{A}$ is strongly decomposable.

We next prove the uniqueness of the closed subalgebra $\mathfrak{B}$. For $a \in \mathfrak{A}$ let T_a be the rgeular representation of $\mathfrak{A}$ in the algebra of bounded operators on $\mathfrak{A}$. The corespondence $f \rightarrow T_{\mu(f)}{}^*$ is a bicontinuous isomorphism of $C(\Omega)$ into the algebra of bounded operators on the conjugate space $\mathfrak{A}^*$. Thus in the terminology of Dunford [5, Theorem 18], $T_b{}^*$ for $b \in B$, is a scalar type spectral operator of class $\mathfrak{A}$ in $\mathfrak{A}^*$, and if $a \in \mathfrak{A}$, $T_a{}^* = T_b{}^* + T_r{}^*$ is a spectral operator in $\mathfrak{A}^*$ of class $\mathfrak{A}$ since $T_r{}^*$ is a commuting generalized nilpotent. By an important theorem of Dunford [5, Theorem 8] the splitting of a spectral operator into scalar and commuting generalized nilpotent parts is unique.

Thus if $\mathfrak{A} = \mathfrak{B}_1 \oplus \mathfrak{R}$, where $\mathfrak{B}_1$ is a closed subalgebra different from $\mathfrak{B}$, there will be an element a in $\mathfrak{A}$ for which the spectral operator T_a^* has two distinct splittings. This contradiction completees the proof.

THEOREM 4.2. *Let $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ and assume Ω is totally disconnected. If there is an integer n such that $x^n = 0$,[8] $x \in \mathfrak{R}$, then $P_{\mathfrak{A}}$ is a bounded set. $\mathfrak{A}$ is therefore strongly decomposable and the decomposition is unique.*

Proof. Let $\sigma: \mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ and let ν be the natural homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{R}$. Since $\mathfrak{A}/\mathfrak{R}$ is complete under the norm $\| \nu(x) \| = \inf_{r \in \mathfrak{R}} \| x + r \|$, σ is a homeomorphism. Therefore there is a constant $\alpha > 0$ such that for each $x \in \mathfrak{A}$ we can choose a corresponding $r \in \mathfrak{R}$ for which $\| x + r \| \leqq \alpha \| (\sigma\nu)(x) \|$. In particular, if $p \in P_{\mathfrak{A}}$, there is an element $r \in \mathfrak{R}$ such that $\| p + r \| \leqq \alpha$. Next we note that since $r^n = 0$, $r \in \mathfrak{R}$,

$$(p + r)^{n+k} = \sum_{j=0}^{n-1} C_j^{n+k} p r^j, \qquad k = 0, 1, 2, \cdots.$$

We assert next that there exist scalars $\lambda_0, \cdots, \lambda_{n-1}$ independent of p and r such that

$$p = \lambda_0 (p + r)^n + \lambda_1 (p + r)^{n+1} + \cdots + \lambda_{n-1} (p + r)^{2n-1}.$$

Expanding the right hand side of this expression and collecting terms of like powers of r this just means that $\lambda_0, \cdots, \lambda_{n-1}$ must be solutions of the equation

$$z_0 + z_1 + \cdots + z_{n-1} = 1$$

$$C_k^n z_0 + C_k^{n+1} z_1 + \cdots + C_k^{2n-1} z_{n-1} = 0, \qquad k = 1, 2, \cdots, n - 1.$$

A solution certainly exists since it may be easily verified that the determinant of coefficients is non zero. Next for $p \in P_{\mathfrak{A}}$ choose $r \in \mathfrak{R}$ such that $\| p + r \| \leqq \alpha$. Then if $\lambda_0, \cdots, \lambda_{n-1}$ are solution of the above equations, we have

$$\begin{aligned} \| p \| &= \| \lambda_0 (p + r)^n + \cdots + \lambda_{n-1} (p + r)^{2n-1} \| \\ &\leqq | \lambda_0 | \, \| p + r \|^n + \cdots + | \lambda_{n-1} | \, \| p + r \|^{2n-1} \\ &\leqq \alpha^n \sum_{k=0}^{n-1} | \lambda_k | \, \alpha^k. \end{aligned}$$

Since this bound is independent of p, $P_{\mathfrak{A}}$ is a bounded set. An application of Theorem 2.4 completes the proof. This argument yields the following slightly more general result.

COROLLARY 4.3. *Let $\mathfrak{A}$ be a commutative Banach algebra with nilpotent*

---

[8] An equivalent hypothesis is that $\mathfrak{R}$ be nilpotent. This foolows easily from Newton's identities for the symmetric functions.

radical. Let S be a collection of idempotents of $\mathfrak{A}$, and $\nu(S)$ the natural image in $\mathfrak{A}/\mathfrak{R}$. If $\nu(S)$ is bounded in $\mathfrak{A}/\mathfrak{R}$, then S is bounded in $\mathfrak{A}$.

If few idempotents are present in the algebra $\mathfrak{A}$, the tools for proving $\mathfrak{A}$ decomposable are very meagre. However, if $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ and $\dim \mathfrak{R} < \infty$, then a device due to Arens [1, 2] enables us to reduce the problem to previous results. Specifically, if $\mathfrak{A}$ is a Banach algebra and $\mathfrak{A}^{**}$ is its second conjugate space, Arens defined a multiplication in $\mathfrak{A}^{**}$ under which $\mathfrak{A}^{**}$ is a Banach algebra containing $\mathfrak{A}$ as a closed weak * dense subalgebra under the natural imbedding. The reader should consult [2] for details of this construction. We note one fact which is important in the following. If $x \in \mathfrak{A}^{**}$ and $\{x_\alpha\} \subset \mathfrak{A}$ is a directed family converging to x in the weak * topology, then for each $y \in \mathfrak{A}$ and each $z \in \mathfrak{A}^{**}$, $yx_\alpha \to yx$ and $x_\alpha z \to xz$ in the weak * topology [2, Theorem 3.2]. This clearly shows that if $\mathfrak{A}$ is commutative, $\mathfrak{A}$ is contained in the center of $\mathfrak{A}^{**}$.

In the following discussion we will always consider $\mathfrak{A}^{**}$ as a Banach algebra without making specific mention of the fact. For a continuous linear map ν, we let ν^{**} denote the canonical second adjoint map, and for $\mathfrak{K} \subset \mathfrak{A}$ we let

$$\mathfrak{K}^{\perp} = \{f \in \mathfrak{A}^{*} : f(\mathfrak{K}) = 0\}.$$

LEMMA 4.3. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be Banach algebras, and ν be a continuous homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$. Then ν^{**} is a homomorphism of $\mathfrak{A}^{**}$ into $\mathfrak{B}^{**}$ (with respect to the Arens multiplication). If ν maps $\mathfrak{A}$ onto $\mathfrak{B}$ with kernel $\mathfrak{K}$, then ν^{**} maps $\mathfrak{A}^{**}$ onto $\mathfrak{B}^{**}$ with kernel $\mathfrak{K}^{\perp\perp}$.*

Proof. Note first that ν^{**} is continuous with respect to the weak * topologies in $\mathfrak{A}^{**}$ and $\mathfrak{B}^{**}$. Since $\mathfrak{A}$ is weak * dense in $\mathfrak{A}^{**}$, for $x, y \in A^{**}$ pick nets $\{x_\alpha\}$, $\{y_\beta\}$ in $\mathfrak{A}$ converging, weak *, to x, y respectively. By the appropriate weak * continuity of multiplication in $\mathfrak{A}^{**}$ and $\mathfrak{B}^{**}$ we have for fixed α, $\nu^{**}(x_\alpha)\nu^{**}(y_\beta) \to \nu^{**}(x_\alpha)\nu^{**}(y)$, and $\nu^{**}(x_\alpha y_\beta) \to \nu^{**}(x_\alpha y)$. Since ν^{**} agrees with ν on $\mathfrak{A}$, it follows that $\nu^{**}(x_\alpha y) = \alpha^{**}(x_\alpha)\nu^{**}(y)$. Applying this argument again we have $\nu^{**}(xy) = \nu^{**}(x)\nu^{**}(y)$. If ν maps $\mathfrak{A}$ onto $\mathfrak{B}$ with kernel $\mathfrak{K}$, then ν^{*} is one to one and $\nu^{*}(\mathfrak{B}^{*}) = \mathfrak{K}^{\perp\perp}$. The last two statements now follows easily.

THEOREM 4.4. *Let $\mathfrak{A}$ be a commutative Banach algebra with finite dimensional radical $\mathfrak{R}$. If $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ for a compact Hausdorff space Ω, then $\mathfrak{A}$ is strongly decomposable and the decomposition is unique.*

Furthermore, if $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$, the annihilator of $\mathfrak{R}$ in $\mathfrak{B}$ has deficiency in $\mathfrak{B} \leq \dim \mathfrak{R}$.

Proof. We note first that $C^{**}(\Omega)$ is isometrically isomorphic to $C(\Omega')$ for some extremely disconnected compact Hausdorff space Ω'. This is a consequence of results of Arens and Grothendieck. In [1] Arens showed that $C^{**}(\Omega)$ is isometrically isomorphic to $C(\Omega')$ for a compact space Ω'. But by [9, Theorem 2] if $C(\Omega')$ is an adjoint space, Ω' must be extremely disconnected.

If ν is the natural homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{R}$, $\sigma\colon \mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$, then $\mu = \sigma\nu$ is a continuous homomorphism of $\mathfrak{A}$ onto $C(\Omega)$ with kernel $\mathfrak{R}$. By the previous lemma μ^{**} is a homomorphism of $\mathfrak{A}$ onto $C^{**}(\Omega)$ with kernel $\mathfrak{R}^{\perp\perp}$ and since $\dim \mathfrak{R} < \infty$, $\mathfrak{R} = \mathfrak{R}^{\perp\perp}$. We assert next that $\mathfrak{R}$ is the radical of $\mathfrak{A}^{**}$.

Let $\mathfrak{N}$ be the radical of $\mathfrak{A}^{**}$. If $x \in \mathfrak{N}$ $\| x^n \|^{1/n} \to 0$, and therefore $\| \nu^{**}(x)^n \|^{1/n} \to 0$ implying $\nu^{**}(x) = 0$ or $x \in \mathfrak{R}$. Now the radical of $\mathfrak{A}^{**}$ contains all topologically nilpotent ideals of $\mathfrak{A}^{**}$, if $x \in \mathfrak{R}$, $y \in \mathfrak{A}^{**}$, then

$$\| (xy) \|^{1/n} = \| x^n y^n \|^{1/n} \leqq \| y \| \cdot \| x^n \|^{1/n} \to 0.$$

Therefore the ideal $x\mathfrak{A}^{**}$ is topologically nilpotent, and consequently $x \in \mathfrak{N}$.

Since R is contained in the center of $\mathfrak{A}^{**}$, the idempotents of $\mathfrak{A}^{**}$ commute by Lemma 2.3. Applying Lemma 2.1 it is easy to see that since $P_{C(\Omega')}$ satisfies property (I), $P_{\mathfrak{A}^{**}}$ does also. Therefore, it follows by Corollary 3.3 that $P_{\mathfrak{A}^{**}}$ is a bounded set, and consequently by Theorem 2.4 there exists a unique subalgebra $\mathfrak{B}'$ of $\mathfrak{A}^{**}$, necessarily closed, such that $\mathfrak{A}^{**} = \mathfrak{B}' \oplus \mathfrak{R}$. If $\mathfrak{B} = \mathfrak{B}' \cap \mathfrak{A}$, $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{R}$ since $\mathfrak{R} \subset \mathfrak{A}$. The uniqueness of $\mathfrak{B}$ in $\mathfrak{A}$ is a consequence of Theorem 4.1.

To prove the last statement we adapt a technique of Feldman [6, Theorem 2]. Let $\mathfrak{J} = \{x \in \mathfrak{B} \mid x\mathfrak{R} = 0\}$. $\mathfrak{J}$ is a closed ideal in $\mathfrak{B}$. Therefore $\mu(\mathfrak{J})$ is a closed ideal in $C(\Omega)$, and consequently there exists a closed set F of Ω such that $\mu(\mathfrak{J})$ is the set of all functions in $C(\Omega)$ which vanish on F. We assert F is finite with cardinality $\leqq n$, the dimension of $\mathfrak{R}$. If the cardinality of F exceeds n, there exist elements $x_1, \cdots, _{n+1}$, $y_1, \cdots, y_{n+1} \in \mathfrak{B}$, $y_i x_i = x_i$, $i = 1, \cdots, n+1$; $y_i y_j = 0$, $i \neq j$; and $x_i \mathfrak{R} \neq \{0\}$, $i = 1, \cdots, n+1$. Therefore, for each index i pick $r_i \in R$ such that $x_i r_i \neq 1$. Since $\dim R = n$, $x_1 r_1, \cdots, x_{n+1} r_{n+1}$ are linearly dependent. Therefore for some i

$$x_i r_i = \sum_{\substack{j=1 \\ i \neq j}}^{n+1} \lambda_j x_j r_j.$$

But then

$$x_i r_i = y_i x_i r_i = \sum_{\substack{j=1 \\ i \neq j}}^{n+1} \lambda_j y_i x_j r_j = 0$$

which is a contradiction.

Clearly the deficiency of $\mathfrak{J}$ equals the cardinality of F since F is a finite set. This completes the proof.

V. By Theorem 4.2 if $\mathfrak{A}/\mathfrak{R} = C(\Omega)$, Ω is totally disconnected, and $\mathfrak{R}^n = 0$, then $\mathfrak{A}$ is (strongly) decomposable. In this section we shall show that this theorem is false if either of these hypotheses are dropped. We first show that $\mathfrak{A}/\mathfrak{R} \simeq C(\Omega)$ is not enough to guarantee the decomposability of $\mathfrak{A}$. Our construction depends on the following result.

LEMMA 5.1. *There exists a commutative n-dimensional Banach algebra* $\mathfrak{A}_n$ *with radical* $\mathfrak{R}_n$ *having the following properties:*

(1) $\mathfrak{A}_n = \{\lambda p_n\} \oplus \mathfrak{R}_n$, $p_n^2 = p_n$,

(2) $\| p_n \| = 2^n$, $\inf_{r \in \mathfrak{R}_n} \| p_n + r \| = 2$.

Proof. Let $\{e_0, \cdots, e_{n-1}\}$ be a basis for an n-dimensional vector space $\mathfrak{A}_n$ over the complex field. For multiplication we define $e_0^2 = e_0$, $e_0 e_j = e_j e_0 = 0$, $j \neq 0$. For $j, k \geqq 1$,

$$e_j e_k = e_{j+k} \quad \text{if} \quad j + k \leqq n-1$$
$$= 0 \quad \text{otherwise.}$$

Clearly $\mathfrak{A}_n$ is a commutative associative algebra, and if $p_n = e_0$, (1) follows. Now let $r = e_1$, then $r^k = e_k$, $1 \leqq k < n$, and $r^k = 0$, $k \geqq n$. Clearly $r, r^2, \cdots, r^{n-1}$ span $\mathfrak{R}_n$. If we let $f = p_n/2 + r$, then for $k < n$, $f^k = p_n/2^k + r^k$, and for $k > n$, $f^k = p_n/2^k$. Clearly $f, \cdots, f^k$ are independent and span $\mathfrak{A}_n$. Therefore each x in $\mathfrak{A}_n$ may be written $x = \sum_{k=1}^{n} \alpha_k f^k$, and we define $\| x \| = \sum_{k=1}^{n} | \alpha_k |$. Under this norm $\mathfrak{A}_n$ is obviously a Banach space.

To show $\| xy \| \leqq \| x \| \| y \|$ for $x, y \in \mathfrak{A}_n$ observe first that $f^k f^j = f^{k+j}$ if $k + j \leqq n$, and $f^k f^j = 2^{-(k+j-n)} f^n$ if $k + j > n$. Therefore if

$$y = \alpha_1 f + \cdots + \alpha_n f^n,$$
$$f^k y = \alpha_1 f^{k+1} + \cdots + \alpha_{n-k-1} f^{n-1} + (\alpha_{n-k} + 2^{-1}\alpha_{n-k-1} + \cdots + 2^{-k}\alpha_n) f^n.$$

Therefore

$$\| f^k y \| = | \alpha_1 | + \cdots + | \alpha_{n-k-1} | + | \alpha_{n-k} + 2^{-1}\alpha_{n-k+1}$$
$$+ \cdots + 2^{-k}\alpha_n | \leqq | \alpha_1 | + \cdots + | \alpha_n | = \| y \|.$$

Hence if $x = \beta_1 f + \cdots + \beta_n f^n$, $\| xy \| < | \beta_1 | \| fy \| + \cdots + | \beta_n | \| f^n y \| \leqq \| x \| \| y \|$. Lastly, since $f^n = 2^{-n} p_n$ and $\| f \| = 1$ we have $\| p_n \| = 2^n$. Also, since $\| f \| = \| p_n/2 + r \| = 1$, $\inf_{r \in \mathfrak{R}_n} \| p_n + r \| \leqq 2$. But if there exists $t \in \mathfrak{R}_n$ such that $\| p_n + t \| < 2$, then $\| p_n/2 + t/2 \| < 1$ and $\| (p_n/2 + t/2)^n \| < 1$. However, $p_n t = 0$ for $t \in \mathfrak{R}$. Therefore $(p_n/2 + t/2)^n = p_n/2^n$, and $\| p_n \| < 2^n$. This is a contradiction.

We now define $\mathfrak{A}_0 = \{\{x_n\} : x_n \in \mathfrak{A}_n, \lim_{n\to\infty} \| x_n \| = 0\}$. If the algebraic operations are defined coordinate-wise and for $x \in A_0$, $\| x \| = \sup_n \| x_n \|$, then $\mathfrak{A}_0^{\bullet}$ is a commutative Banach algebra. Let $\mathfrak{A}$ be the algebra obtained by adjoining an identity e to $\mathfrak{A}_0$ in the usual way. For $y = \alpha e + x$, $x \in \mathfrak{A}_0$, $\| y \| = | \alpha | + \| x \|$. We note that if q is an idempotent of $\mathfrak{A}$, q or $e - q$ belongs to $\mathfrak{A}_0$, since for any idempotent in $\mathfrak{A}_0$ all but a finite number of coordinates are zero. If $x \in \mathfrak{A}_n$, then the natural isometric image of x in $\mathfrak{A}$ will be denoted by x_n. Since $\| p_n \| = 2^n$, and $\| q \| \geqq \| p \|$ if $qp = p$, it is clear that any uniformly bounded family of orthogonal idempotents in $\mathfrak{A}$ is finite.

THEOREM 5.2. *If $\mathfrak{R}$ is the radical of $\mathfrak{A}$, then $\mathfrak{A}/\mathfrak{R} \cong (c)$, the ring of convergent sequences, and $\mathfrak{A}$ is not decomposable.*

Proof. We verify first that $\Phi_{\mathfrak{A}_0}$ is totally disconnected. If $\mathfrak{A}' = \{x \in \mathfrak{A} : x_n = 0, n \text{ sufficiently large}\}$, then $\mathfrak{A}'$ is dense in $\mathfrak{A}_0$, and any $x \in \mathfrak{A}'$ may be written $x = \sum_{k=1}^{n} \lambda_k p_k + r$, $r \in \mathfrak{R}$. If ϕ, ϕ' are two distinct points in $\Phi_{\mathfrak{A}_0}$, there exists an $x \in \mathfrak{A}'$ for which $\phi(x) \neq \phi'(x)$. In fact for some k, $\phi(p_k) \neq \phi'(p_k)$, which shows that $\Phi_{\mathfrak{A}_0}$ is totally disconnected.

For $x \in \mathfrak{A}$ let $\bar{x}$ be the corresponding element in $\mathfrak{A}/\mathfrak{R}$. If $\bar{p}$ is any idempotent of $\mathfrak{A}/\mathfrak{R}$ then $\| \bar{p} \| \leqq 3$. For by Lemma 2.1 there is a unique idempotent $q \in \mathfrak{A}$ such that $\bar{q} = \bar{p}$. Hence, assuming $q \neq 0$, if $q \in \mathfrak{A}_0$, $\| \bar{p} \| = \inf_{r \in R} \| q + r \| = 2$ if $e - q \in \mathfrak{A}_0$, $\| e - \bar{p} \| = 2$ and $\| \bar{p} \| \leqq 3$. An application of Theorem 2.4 shows that the span of the idempotents of $\mathfrak{A}/\mathfrak{R}$ is dense in $\mathfrak{A}/\mathfrak{R}$. It is now easily verified that for $\xi = \{\xi_k\} \in (c)$, if $\xi_0 = \lim_{k\to\infty} \xi_k$ then the mapping $\xi \to \xi_0 \bar{e} + \sum_{k=1}^{\infty} (\xi_k - \xi_0) \bar{p}_k$ is a bicontinuous isomorphism of (c) onto $\mathfrak{A}/\mathfrak{R}$.

If $\mathfrak{A}$ were decomposable, then there would exist an isomorphism ν of (c) into $\mathfrak{A}$. But if $\{e_i\}$ is any infinite collection of orthogonal idempotents of (c), by [4, Corollary 2.2] $\{\nu(e_i)\}$ is a bounded set in $\mathfrak{A}$. Any such family is necessarily finite, hence no such isomorphism can exist. $\mathfrak{A}$ is therefore not decomposable.

Actually, the following stronger result is true. If σ is any homomorphism of (c) into $\mathfrak{A}$, then σ has finite dimensional range. A sketch of the proof is as follows. By [4, Theorem 4.3] we may write $\sigma = \mu + \lambda$, where μ is continuous and λ maps (c) into the radical of $\mathfrak{A}$. The kernel $\mathfrak{K}$ of μ is the set of convergent sequences vanishing on some closed set F of $\Phi_{(c)}$. The latter is of course just the one point compactification of the positive integers.

Using the continuity of μ and the fact that any bounded family of orthogonal idempotents in $\mathfrak{A}$ is finite, it is easy to see that $\mu(z)=0$ if infinitely many of the coordinates of z are different from zero. Therefore, F is a finite set of points, and since $(c)/\mathfrak{R}\simeq C(F)$, μ has finite dimensional range. To show $\lambda\equiv 0$, one observes first that λ is a homomorphism on (c_0). Since for any $z\in(c_0)$, $[\lambda(z)^n]_n=0$, we have for $z\geqq 0$, $z\in(c_0)$, $\lambda(z)_n=[\lambda(z^{1/n})^n]_n=0$. Consequently, $\lambda\equiv 0$.

Next we show that $\mathfrak{R}^n=0$ is also not enough to guarantee decomposability. The example is similar to the previous one.

LEMMA 5.3. *For each integer n there exists a three dimensional commutative Banach algebra $\mathfrak{A}_n$ with two dimensional radical $\mathfrak{R}_n$ having the following properties:*

(1) $\mathfrak{A}_n=\{\lambda p_n\}\oplus\mathfrak{R}_n,\ p_n^2=p_n,$

(2) $\|p_n\|=n^3;\ \inf_{r\in\mathfrak{R}_n}\|p_n+r\|=n.$

Proof. Let $\mathfrak{A}_n$ be the three dimensional commutative algebra generated by elements p_n, r, s, where $p_n^2=p_n$; $p_nr=rp_n=0$; $r^2=s$ and $rs=0$. Now elements $f_1=p/n+r$, $f_2=p/n^2+s$, $f_3=p/r^3$ also form a basis for $\mathfrak{A}_n$. Therefore, if $x\in\mathfrak{A}_n$ and $x=\sum_{i=1}^{3}\alpha_if_i$, we define $\|x\|=\sum_{i=1}^{3}|\alpha_i|$. We leave to the reader the verification that $\mathfrak{A}_n$ is a Banach algebra under this norm. Assertaions (1) and (2) are clear.

We now define $\mathfrak{A}=l_1(\mathfrak{A}_n)=\{\{x_n\}: x_n\in\mathfrak{A}_n,\ \sum_{n=1}^{\infty}\|x_n\|<\infty\}$. The algebraic operations are defined pointwise and for $x\in\mathfrak{A}$, $\|x\|=\sum_{n=1}^{\infty}\|x_n\|$.

THEOREM 5.4. *If $\mathfrak{R}$ is the radical of the above algebra $\mathfrak{A}$, $\mathfrak{R}^3=\{0\}$, and $\mathfrak{A}$ is not decomposable.*

Proof. Clearly $\mathfrak{R}=\{x\in\mathfrak{A}: x_n\in\mathfrak{R}_n\}$, and since $\mathfrak{R}_n^3=\{0\}$, it follows that $\mathfrak{R}^3=\{0\}$. Let $\bar{p}_n\in\mathfrak{A}/\mathfrak{R}$ be the images of the primitive idempotents p_n. By the construction $\|\bar{p}_n\|=\inf_{r\in\mathfrak{R}}\|p_n+r\|=\inf_{r\in\mathfrak{R}_n}\|p_n+r\|=n$. If there exists an isomorphism μ of $\mathfrak{A}/\mathfrak{R}$ into $\mathfrak{A}$, then by [4, Corollary 2.2] there exists a constant M such that

$$\|\mu(\bar{p}_n)\|\leqq M\|\bar{p}_n\|^2=Mn^2, \qquad n=1,2,\cdots.$$

But by the construction it is clear that for each n

$$\sup_{k=1,\cdots,n}\|\mu(\bar{p}_k)\|\geqq n^3.$$

These two facts are obviously incomparable, hence the theorem follows.

In the above example, $\mathfrak{A}/\mathfrak{R}$ is isometrically isomorphic with the algebra $\mathfrak{B}$ of sequences $\{\alpha_n\}$ such that $\sum_{n=1}^{\infty} n \mid \alpha_n \mid < \infty$. For if $\bar{x} = \sum_{n=1}^{N} \alpha_n \bar{p}_n$, then

$$\| \bar{x} \| = \inf_{r \in \mathfrak{R}} \| \sum_{n=1}^{N} \alpha_n p_n + r \| = \inf_{r \in \mathfrak{R}} \sum_{n=1}^{N} \| \lambda_n p_n + r_n \|$$
$$= \sum_{n=1}^{N} \inf_{r_n \in \mathfrak{R}_n} \| \alpha_n p_n + r_n \| = \sum_{n=1}^{N} n \mid \alpha_n \mid .$$

Thus we have an alegbraic isometry of a dense subalgebra of $\mathfrak{B}$ onto a dense subalgebra of $\mathfrak{A}/\mathfrak{R}$ which can be extended to an isometry of $\mathfrak{B}$ onto $\mathfrak{A}/\mathfrak{R}$. The following result shows that for algebras for which $\mathfrak{A}/\mathfrak{R} \simeq l_1$ and $\mathfrak{R}^n = \{0\}$ no such construction as the above is possible. We are assuming coordinatewise multiplication in l_1.

THEOREM 5.5. *Let A be a commutative Banach algebra such that $A/R \simeq l_1$. Then if $\mathfrak{R}^n = \{0\}$ for some integer n, $\mathfrak{A}$ is strongly decomposable, and the strong decomposition is unique.*

In the light of Corollary 4.3 this will follow from the following:

THEOREM 5.6. *Let $\mathfrak{A}/\mathfrak{R} \simeq l_1$. Then $\mathfrak{A}$ is decomposable if and only if the primitive idempotents of $\mathfrak{A}$ form a bounded set. If the primitive idempotents of $\mathfrak{A}$ are bounded, then $\mathfrak{A}$ is strongly decomposable, and the strong decomposition is unique.*

Proof. Suppose $\mathfrak{A}$ is decomposable. Then there exists an isomorphism μ of l_1 into $\mathfrak{A}$ which carries the set S of primitive idempotents of l_1 onto the set T of primitive idempotents of $\mathfrak{A}$. Since S is a bounded set in l_1, by [4, Corollary 2.2] $\mu(S) = T$ is a bounded set in $\mathfrak{A}$. Conversely, suppose T is a bounded set in $\mathfrak{A}$. Let ν be the natural homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{R}$, λ the isomorphism of $\mathfrak{A}/\mathfrak{R}$ onto l_1, and let $\mu = \lambda\nu$. For each primitive idempotent $e_i \in l_1$, there exists a unique primitive idempotent $f_i \in \mathfrak{A}$ such that $\mu(f_i) = e_i$. Let $\mathfrak{B} = \{x \in \mathfrak{A} \mid x = \lim_{n \to \infty} \sum_{k=1}^{n} \alpha_k f_k$, where $\sum_{k=1}^{\infty} \mid \alpha_k \mid < \infty\}$. By the boundedness of $\{f_k\}$, $\sum_{k=1}^{\infty} \alpha_k f_k$ converges absolutely. Therefore since μ is continuous, $\mu(x) = \sum_{k=1}^{\infty} \alpha_k e_k$, and μ maps $\mathfrak{B}$ in a one to one fashion onto l_1. For $x \in \mathfrak{B}$, $\| x \| \leqq \sum_{k=1}^{\infty} \mid \alpha_k \mid \| f_k \| \leqq M \| \mu(x) \|$; therefore $\mathfrak{B}$ is closed, and $\mathfrak{A}$ is strongly decomposable. The uniqueness of this strong decomposition can now be verified easily by adapting the method of Theorem 2.4.

We note that under the hypothesis of Theorem 5.5 the decomposition

will not in general be unique. To see this we define $\mathfrak{A} = l_1 \oplus \{\alpha r\}$, where $r^2 = r(l_1) = 0$. Let μ be the natural isometric injection of l_1 into $\mathfrak{A}$. If f is a discontinuous linear functional on l_1 which vanishes on $(l_1)^2$, then $\lambda(z) = \mu(z) + f(z)r$, for $z \in l_1$ defines an isomorphism of l_1 into $\mathfrak{A}$. If $\mathfrak{B}' = \lambda(l_1)$, then it is clear that $\mathfrak{A} = \mathfrak{B}' \oplus \mathfrak{R}$, and $\mathfrak{B}'$ is not closed in $\mathfrak{A}$.

UNIVERSITY OF CALIFORNIA, BERKELEY,
UNIVERSITY OF CALIFORNIA, LOS ANGELES.

REFERENCES.

[1] R. Arens, "Operations induced in functions classes," *Monatshefte für Mathematik*, vol. 55 (1951), pp. 1-19.

[2] ———, "The adjoint of a bilinear operation," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 839-848.

[3] W. G. Bade, "On Boolean algebras of projections and algebras of operators," *Transactions of the American Mathematical Society*, vol. 80 (1955), pp. 345-360.

[4] ——— and P. C. Curtis, Jr., "Homomorphisms of commutative Banach algebras," *American Journal of Mathematics*, vol. 82 (1960), pp. 589-608.

[5] N. Dunford, "Spectral operators," *Pacific Journal of Mathematics*, vol. 4 (1954), pp. 321-354.

[6] C. Feldman, "The Wedderburn principal theorem in Banach algebras," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 771-777.

[7] L. Gillman and M. Henriksen, "Rings of continuous functions in which every finitely generated ideal is principal," *Transactions of the American Mathematical Society*, vol. 82 (1956), pp. 366-391.

[8] G. Glaeser, "Étude de certaines algèbres tay loriennes," *Journal d'Analyse Mathematique*, vol. 6 (1958), pp. 1-124.

[9] A. Grothendieck, "Une caractérisation des espaces L'," *Canadian Journal of Mathematics*, vol. 7 (1955), pp. 552-561.

[10] M. Henriksen, "Some remarks on a paper of Aronszajn and Panitchpakdi," *Pacific Journal of Mathematics*, vol. 7 (1957), pp. 1619-1621.

[11] I. Kaplansky, "Normed algebras," *Duke Mathematical Journal*, vol. 16 (1949), pp. 399-418.

[12] H. A. Kober, "A theorem on Banach spaces," *Compositio Mathematica*, vol. 7 (1939), pp. 135-140.

[13] L. H. Loomis, *An introduction to abstract harmonic analysis*, New York, 1953.

[14] C. E. Rickart, "On spectral permanence for certain Banach algebras," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 191-196.

[15] ———, *General theory of Banach algebras*, New York, 1960.

[16] G. E. Silov, "On a decomposition of a commutative normed ring in a direct sum of ideals," *Mathematičeskiĭ Sbornik N. S.*, vol. 32 (1954), pp. 353-364.

[17] D. Zelinsky, "Raising idempotents," *Duke Mathematical Journal*, vol. 21 (1954). pp. 315-322.

FUNCTIONS THAT ARE HARMONIC OR ZERO.* [1]

By JOHN W. GREEN.

1. **Introduction.** In a paper on integral characterizations of harmonic functions, E. F. Beckenbach [1] considered the following proposition:

THEOREM 1. *Let $u(x,y)$ be of class C' in a domain D and have the property that at each point (x_0,y_0) of D either $u(x_0,y_0)=0$ or u is harmonic in some neighborhood of (x_0,y_0). Then u is harmonic in D.*

If u is only of class C^0, the theorem is false, as the example $u=|x|$ shows. If u is of class C'', the result is trivial, since it follows immediately that $\nabla^2 u\equiv 0$. In [1] is given a proof of the theorem under the additional hypothesis that u_{xx} and u_{yy} are summable.[2] In the present paper, Theorem 1 is proved without any addditional assumptions on u, applications to potential theory made, and an analogous theorem proved for analytic functions.

2. **A preliminary lemma.** Let E be the set of points of D at which u is not harmonic. The set E is closed relative to D and $u=0$ on E. We prove the following lemma:

LEMMA 1. $u_x=u_y=0$ *on* E.

Let (x_0,y_0) be a point of E at which $u_y\neq 0$. Now E is contained in the locus of the equation $u(x,y)=0$; hence all points of E in some neighborhood of (x_0,y_0) will lie on the curve $y=f(x)$ of class C' which is the solution of the implicit equation $u(x,f(x))=0$. The function u is harmonic above and below this curve and continuous together with u_x and u_y across it. Conjugate functions $v+$ and $v-$ to u above and below the curve can be constructed by the usual integral formula $\int(\partial u/\partial n)\,ds$. Since integrals along the curves are allowed, $v+$ and $v-$ differ on the curve by a constant, which can be taken to be zero. By Painleve's theorem on the continuation of an analytic function

* Received March 7, 1960.

[1] The preparation of this paper was sponsored, in part, by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] It is perhaps not completely clear in [1] that what is needed and used there is that u_x be absolutely continuous with respect to x for almost all y, that u_{xx} be summable over D, and the same holds with letters interchanged.

across a rectifiable curve on which it remains continuous, one sees that u continues harmonically into itself across the curve. Therefore (x_0, y_0) does not belong to E after all. The case of u_x is handled analogously.

3. A special case of Theorem 1. We now prove Theorem 1 in the case D is a circle and u is harmonic on the boundary of D; in this case $\bar{E} \subset D$.

LEMMA 2. *Let D' be that component of $D - E$ which extends to the boundary of D. Then u_{xx} and u_{yy} are summable over D'.*

Let w be u_x or u_y. In either case w is harmonic in D' and on the outer boundary of D'; it is continuous and equal to zero on the inner boundary, which is composed of points of E. It is clearly possible to define a function in D' with finite Dirichlet integral and with the same boundary values as w. Since the harmonic function with prescribed continuous boundary values, when it exists, minimizes the Dirichlet integral, w has finite Dirichlet integral over D'. Thus $|w_x|^2$ and $|w_y|^2$ are summable over D' and this implies that u_{xx} and u_{yy} are.

LEMMA 3. *Let f be a continuous function on (a, b) and F a closed set in (a, b). Let $f(x) = 0$ for x in F, and let f be absolutely continuous on each of the open intervals of which the set $(a, b) - F$ is formed. Finally let f' be summable on $(a, b) - F$. Then f is absolutely continuous on (a, b) and $f' = 0$ almost everywhere on F.*

Let x_0 be a point of F and set $g(x) = f'(x)$ for x in $(a, b) - F$ and $g(x) = 0$ for x in F. Define

$$h(x) = \int_{x_0}^{x} g(t)\, dt.$$

If $x \in F$, then $h(x) = \sum \int_{\alpha_i} f'(t)\, dt$, where α_i are the intervals of $(a, b) - F$ lying between x_0 and x. But if $\alpha_i = (a_i, b_i)$,

$$\int_{\alpha_i} f'(t)\, dt = f(b_i) - f(a_i) = 0.$$

Thus $h(x) = 0$ for x in F. If $x \in (a, b) - F$, let x_1 be one of the ends of the interval of $(a, b) - F$ in which x lies. Then

$$h(x) = \int_{x_0}^{x_1} g(t)\, dt + \int_{x_1}^{x} g(t)\, dt = 0 + \int_{x_1}^{x} f'(t)\, dt = f(x) - f(x_1) = f(x).$$

We see now that $f \equiv h$ and f is absolutely continuous. Since $f' = g$ almost everywhere, $f' = 0$ almost everywhere in F.

LEMMA 4. *For almost all y, $u_x(x,y)$ is absolutely continuous with respect to x, u_{xx} is summable over D, and $u_{xx}=0$ almost everywhere in $D-D'$.*

Let L_y represent the horizontal line with ordinate y. Since u_{xx} is summable over D', it is summable over $L_y \cdot D'$ for almost all y. For such a y we apply Lemma 3 to u_x, taking $(a,b)-F$ to be $L_y \cdot D'$. The set F consists of points of E, at which $u_x=0$, and of points of other components than D' of $D-E$. But such components would have as boundary points only points of E. Therefore $u\equiv 0$ in such components and $u_x=0$ on F. From Lemma 3 it follows that u_x is absolutely continuous in x, and $u_{xx}=0$ almost everywhere on F. By Fubini's theorem, $u_{xx}=0$ almost everywhere (two dimensional measure) on $D-D'$, and u_{xx} is summable over D.

The special case of Theorem 1 is at hand, since Beckenbach's result is applicable. However, for completeness, the following proof, slightly different from that in [1], is included.

It suffices to prove that $\int (\partial u/\partial n)\,ds=0$ around any rectangle in D with horizontal and vertical sides. For if this is so, one can define a conjugate function

$$v=\int_{(x_0,y_0)}^{(x,y)} (\partial u/\partial n)\,ds,$$

restricting the integration to a horizontal followed by a vertical segment, or vice-versa. Then u and v are easily seen to satisfy the Cauchy-Riemann equations.

Let then (x_0,y_0) and (x_1,y_1) be lower left and upper right vertices of a rectangle R in D. For almost all y

$$\int_{x_0}^{x_1} u_{xx}(x,y)\,dx=u_x(x_1,y)-u_x(x_0,y),$$

and so

$$\int\int_R u_{xx}\,dxdy=\int_{y_0}^{y_1} u_x(x_1,y)\,dy-\int_{y_0}^{y_1} u_x(x_0,y)\,dy.$$

If the corresponding equation with x and y interchanged is put down and added to this one, the right hand side of the resulting equation is the integral of $\partial u/\partial n$ around R. The left hand side is

$$\int\int_R (u_{xx}+u_{yy})\,dxdy=\int\int_{D'\cdot R}(u_{xx}+u_{yy})\,dxdy=0.$$

4. A lemma concerning analytic functions.

LEMMA 5. *Let K be a simply connected domain lying in the open unit*

circle D. Let f be analytic in K and continuous on $\bar{K}$. Let $f=0$ at all boundary points of K lying in D. Then if K has a boundary point in D, $f\equiv 0$ in K.

Let A be the set of boundary points of K at which $f=0$, and B the remaining boundary points of K, at which $f\neq 0$. By the continuity of f, B is a relatively open set on the unit circle $|z|=1$ and all points of B are accessible from K by straight line segments. Let

$$z=\phi(w)$$

map K conformally on the unit circle $|w|<1$. By known properties of the mapping function, B is carried into a relatively open set G on $|w|=1$ in a one-to-one fashion. If F is the complement of G on $|w|=1$, then A is carried into F. By this is meant that if a sequence w_i tends to a point of F, all accumulation points of $\phi(w_i)$ lie on A, and vice-versa.

Suppose K has a boundary point in D. Then F is of positive measure on $|w|=1$. This follows from a theorem of A. Ostrowski [2] which states that in the mapping of a simply connected domain on the unit circle, every neighborhood of any boundary point contains a set of accessible boundary points whose image on $|w|=1$ has positive measure.

Consider the function

$$g(w)=f(\phi(w)).$$

It is analytic and bounded in $|w|<1$, and tends to zero when w tends to a point of F. By a theorem of F. and M. Riesz (see for example [3], p. 156) $f\equiv 0$.

5. Proof of Theorem 1. We return to Theorem 1 in the general case. Let then u be of class C' in a domain D, and u be zero or harmonic at every point of D. Let P be a point of D at which u is not harmonic, and let D' be an open circle about P such that $\bar{D}'\subset D$. Let K be any connected component of $D'-E$.

We first note that K must be simply connected. If it were not, some simple closed curve lying in K would enclose points of E. By mapping the interior of this curve on the unit circle, we would have the situation covered in Section 3 and would conclude that there could be no points of E inside the curve after all.

Consider now the function

$$f(z)=u_x(x,y)-iu_y(x,y).$$

It is analytic in K and continuous in $\bar{K}$; furthermore it is zero at all boundary points of K lying in D', such boundary points being points of E. Also, since K is not all of D', there must be such boundary points. Therefore by Lemma 5, $f \equiv 0$ in K, and so $u_x \equiv u_y \equiv 0$ in K. Since $u = 0$ on the aforementioned boundary points of K, $u \equiv 0$ in K.

It follows that $u \equiv 0$ in $D' - E$, and since $u = 0$ on E, $u \equiv 0$ in D'. Thus u is harmonic in D' and P cannot be a point of E. This concludes the proof.

6. Applications. In [1] it is shown that if u is of class C' in D and if the Dirichlet integral of u over each circle of radius r differs from the integral of $u\partial u/\partial n$ around the perimeter by $o(r^2)$, then u is harmonic in D provided either (a) u does not vanish in D, or (b) u satisfies the conditions stated in the footnote to § 1. These additional conditions arose because of the fact that in the course of the proof, it was shown that u was either zero or harmonic at each point of D. It is clear then that we have the following theorem:

THEOREM 2. *If u is of class C' in D and if for each circle K in D with perimeter C we have*

$$\int\int_K (u_x^2 + u_y^2)\,dxdy - \int_C u(\partial u/\partial n)\,ds = o(r^2),$$

then u is harmonic in D.

Let E be a bounded point set which has a conductor potential in the strict sense; that is, a function u continuous in the finite plane, harmonic except on E, equal to 1 on E, and with the appropriate logarithmic singularity at ∞. If u had continuous partial derivatives, Theorem 1 could be applied to this function $u - 1$ and u would be harmonic in the finite plane, which is impossible. Thus the following theorem may be stated.

THEOREM 3. *No conductor potential in the strict sense can have continuous partial derivatives.*

This is perhaps not surprising. The usual situation for smooth sets is for the conductor distributions of smooth sets to lie on curves, and for the discontinuity of the normal derivative across these curves to be proportional to the linear density. However for general point sets the result is not particularly obvious.

7. A theorem on analytic functions. We can prove a theorem analogous to Theorem 1 for analytic functions. In this case, however, the continuity of the derivatives is not required.

THEOREM 4. *Let $f(z)$ be continuous in a domain D, and at each point of D let f be either analytic or zero. Then f is analytic in D.*

The proof is very similar to that of Theorem 1 and details will be omitted. If $f = u + iv$, then in the analogue of Lemma 2, one proves that u_x, u_y are summable over D'. In the analogue of Lemma 4, one shows that u is absolutely continuous with respect to x and $u_x = 0$ almost everywhere in $D - D'$. To prove the special case of Theorem 4, it is shown that the integral of f around rectangles is zero, and Morera's theorem applies. In proving the general case of the theorem, Lemma 5 is applied directly to f.

Added in Proof. Professor G. R. MacLane has recently pointed out that Theorem 4 is known, and indeed is known as the Radó-Behnke-Stein-Cartan theorem. The contribution of Radó (*Über eine nicht fortsetzbare Riemannsche Mannigfaltigkeit,* Mathematische Zeitschrift, vol. 20 (1924), pp. 1-6) is a proof of our Lemma 5. The corresponding result for functions of several complex variables was proved by P. Thullen (*Uber die wesentlichen Singularitaten, usw.,* Mathematische Annalen, vol. 111 (1935), pp. 137-157). H. Behnke and K. Stein (*Modification komplexer Mannigfaltigkeiten und Riemanncher Gebiete,* Mathematische Annalen, vol. 124 (1951), pp. 1-16) proved our Theorem 4 for several complex variables. This amounts essentially to removing the hypothesis of simple connectedness from Lemma 5. Later and simpler proofs were given by H. Cartan (*Sur une extension d'un Théorème de Radó,* Mathematische Annalen, vol. 125 (1952-53), pp. 49-50) and E. Heinz (*Ein elementarer Beweis des Satzes von Radó-Behnke-Stein-Cartan,* Mathematische Annalen, vol. 131 (1956), pp. 258-259), using a variety of methods. The present method appears to be different from the previous one, and also yields the corresponding theorem for harmonic functions. Of course, the theorem for harmonic functions also follows from the theorem for analytic functions with the aid of our preliminary lemma.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES.

REFERENCES.

[1] E. F. Beckenbach, "On characteristic properties of harmonic functions," *Proceedings of the American Mathematical Society,* vol. 3 (1952), pp. 765-769.

[2] A. Ostrowski, "Quasianalytische Funktionen und asymptotische Entwickelung," *Acta Mathematica,* vol. 53 (1929), p. 251.

[3] L. Bieberbach, *Lehrbuch der Funktionentheorie,* Leipzig, 1934.

FUNDAMENTAL GROUPS ON A LORENTZ MANIFOLD.* †

By J. WOLFGANG SMITH.

1. **Introduction.** Let $\mathfrak{V}$ denote a differentiable manifold of dimension $n \geqq 2$ and class C^p, $2 < p \leqq \infty$. A *metric Lorentz structure* $\mathcal{L}$ on $\mathfrak{V}$ is a second order symmetric covariant tensor field of class C^q, $2 \leqq q \leqq p$, and signature $(2-n)$. Thus $\mathcal{L}$ is a quadratic form having 1 positive and $(n-1)$ negative eigenvalues everywhere on $\mathfrak{V}$. It defines an inner product in the tangent vector spaces to $\mathfrak{V}$ and enables us to distinguish *time-like, null* and *space-like* tangent vectors, according as their length squared is positive, zero or negative, respectively. The totality of null-vectors in a given tangent vector space constitutes a cone with vertex at the origin, sometimes referred to as a *light cone.* If one keeps the light cones but discards the metric tensor field, one obtains what may be called a *conformal Lorentz structure* L on $\mathfrak{V}$.[1] We are interested in studying the *global character* of such a structure. The term *global character* has of course no precise meaning until one indicates a method whereby it may be measured or described. To give an example of one such method, we propose to regard L as a cross-section of an appropriate tensor bundle over $\mathfrak{V}$. The homotopy class of the cross-section measures then a global characteristic of L.[2] In contradistinction to these purely *topological* properties, we shall be concerned with more refined global characteristics of L which are *not* invariant under arbitrary homotopies of the cross-section. Such characteristics may be called *conformal global properties* (c.g.p.) of the Lorentz structure. Perhaps the simplest example of a c.g.p. is furnished by the *barriers* defined in [7].

In the present paper we shall study certain c.g.p. which are described by means of an infinite collection of groups. The basic process employed to

* Received March 21, 1960.

† The results established in this paper have been announced under the same title in *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 46 (1960), pp. 111-114. This research was supported in part by the United States Air Force under contract No. AF-49(638)-42, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[1] $\mathcal{L}$ and L may be referred to as *associated* structures.

[2] This is essentially what the physicists Finkelstein and Misner [3] have done to obtain particle numbers for certain geons.

define these groups has been outlined in [8] and may be summarized as follows. Let x be a point of $\mathfrak{V}$ and X a set of loops in $\mathfrak{V}$ based at x. A homotopy $F: I \times I \rightarrow \mathfrak{V}$ (where I denotes the closed unit interval) is said to be *in* X if the maps $F_t: I \rightarrow \mathfrak{V}$, defined by the restriction of F to $I \times t$, belong to X for all $t \in I$. Two elements $f_0, f_1 \in X$ are *homotopic in* X (written $f_0 \sim f_1$ in X) if there exists a homotopy F in X with $F_i = f_i$ for $i = 0, 1$. Let ξ denote the set of equivalence classes in X under homotopy in X. In the event that a given multiplication in X induces a group structure on ξ we will say that ξ *arises from* X. Apart from the fundamental group $\pi_1(\mathfrak{V}, x)$, numerous other groups arise from suitable loop spaces X. The groups τ_q in question have been obtained in [8] from certain loop spaces T_q, $q = 0, 1, 2, \cdots$; which depend on L.

We shall henceforth use the term *loop space* to denote a set of loops having a common base point, together with a law of multiplication, which gives rise to a group. Two loop spaces X_1, X_2 will be called *equivalent* if there exists a loop space X with $X_i \subset X$, such that the inclusion maps I_i: $X_i \rightarrow X$ induce isomorphisms of the corresponding groups. We agree always to identify groups arising from equivalent loop spaces. To obtain the groups τ_q, in particular, one has a choice of various equivalent loop spaces. In the present paper we shall take as our starting point certain loop spaces $T_q(L, x)$, to be defined in Section 2, which permit us to state the proof of basic results with optimum economy. The proof of equivalence for the $T_q(L, x)$ and the T_q considered in [8] is no more than an exercise, and will be omitted.

This work is part of a general effort to gain insight into the global structure of general Lorentz manifolds. Whatever interest the τ-groups may have in themselves, we believe that their significance must ultimately be judged in terms of their ability to contribute towards such an insight. It is too early to predict whether the new groups will be fruitful or not, in this sense. The most that can be claimed for them at this stage is that they bring to light some hitherto unrecognized global properties of a general Lorentz manifold. The present paper is devoted to the proof of four theorems concerning the τ-groups. The first (§ 4) establishes a general connection between the τ-groups and the fundamental group π_1 of $\mathfrak{V}$. The second and third theorems (§ 5) are concerned with the dependence of the new groups on base point, and the last (§ 7) tells essentially what happens when the Lorentz structure is flat and complete.

2. The loop spaces. Our first task is to define the loop spaces $T_q(L, x)$. By a *path* we shall understand a piecewise differentiable map $f: I \rightarrow \mathfrak{V}$, and

we will say that f *starts* at $f(0)$ and *ends* at $f(1)$. If f and g are two paths such that f ends where g starts, then one can define a product fg in the usual way,[3] giving a path which starts at $f(0)$ and ends at $g(1)$. The totality of paths in $\mathfrak{V}$, together with this law of multiplication, will be denoted by P^*. A path f, being piecewise differentiable, determines a unique tangent vector at $f(t)$ for every regular point $t \in I$, and two distinct tangent vectors for every singular $t \in I$. The path will be called *time-like* if all its tangent vectors are time-like. Two time-like vectors belonging to the same tangent vector space are called *oppositely* or *equally directed,* depending on whether they are separated or not by the light cone. A time-like path f is said to have an *interior corner* at $t \in I$ if t is a singular point of f, and if the corresponding tangent vectors to f at $f(t)$ are oppositely directed. It is said to have an *exterior corner* if it is closed, and if its initial and terminal tangent vectors are oppositely directed. We define a *q-loop based at x* (q a nonnegative integer) to be a time-like path which starts and ends at x and has exactly q corners.[4] Products involving q-loops based at x will be the basic ingredients of $T_q(L,x)$. These alone, however, would only give rise to a semi-group, and one must add some additional ingredients in order to obtain an identity and inverse elements. On the other hand, one must not add too much, for this would only lead to the usual fundamental group. We consequently define a *sting based at x* to be a path of the form ff^{-1} with $f(0) = x$. The set generated by q-loops and stings, based at x, does give rise to the desired group $\tau_q(L,x)$. It will be convenient, however, to use an equivalent loop space containing certain additional elements. For this purpose a few more definitions will be required. If f is an arbitrary path and $y \in f(I)$, we define a *y-factorization* of f to be one of the form f_+f_- with $f_+(1) = y$. If g is a path starting at y and $f^* = f_+gg^{-1}f_-$, we will say that f^* is obtained from f by *insertion of a sting.* Similarly, f is obtained from f^* by *deletion of a sting.* We define a *generalized q-loop* to be a path which can be obtained from a q-loop by insertion and deletion of stings (applied a finite number of times.) The constant path e_x (mapping I on x) will also be regarded as a generalized q-loop based at x, for all values of q. One observes that a p-loop f with $p \leqq q$ and p having the same *parity* as q (this means $p+q$ is even) is likewise a generalized q-loop. For clearly f can be turned into a q-loop by insertion of stings. The generalized q-loops are in a sense limits of q-loops, and it will be good to have them in our loop spaces. We consequently define $L_q^*(L,x)$ to be the set of all generalized q-loops based at x, and $T_q^*(L,x)$ to be the

[3] Cf., for instance, P. J. Hilton [4].

[4] This is a slightly weaker definition than the one given in [8].

subset of P^* generated by $L_q^*(L,x)$. In other words, the elements of $T_q^*(L,x)$ are finite products of generalized q-loops based at x. One verifies immediately that $T_q^*(L,x)$ is a loop space.

The asterisk indicates that we are still not satisfied. This stems from the fact that it will be greatly advantageous to deal exclusively with paths whose parametrization has been *standardized* in some way. For this purpose we will put a Riemannian metric[5] $\mathcal{R}$ on $\mathfrak{V}$ and use arc length over total length (relative to $\mathcal{R}$) as standard parameter. One obtains thus a map S from P^* into the set P of all standardized paths. We define a multiplication in P to be the multiplication in P^* followed by S. The multiplication in P is thus associative. We further note the following properties of S which are easily verified:

(i) for all loops $f \in P^*$, $Sf \sim f$ in P^*;

(ii) for all $f \in T_q^*(L,x)$, $Sf \sim f$ in $T_q^*((L,x)$;

(iii) for $f, g \in P^*$ with fg defined, $S(fg) = S(f)S(g)$, the multiplication on the right being understood in P;

(iv) $f \sim g$ in X implies $Sf \sim Sg$ in $S(X)$.

Now let $T_q(L,x) = S(T_q^*(L,x))$. Clearly $T_q(L,x)$ is again a loop space (multiplication being understood in P), and it follows by properties (ii), (iii) and (iv) that $T_q(L,x)$ is equivalent to $T_q^*(L,x)$.[6] We take the sets $T_q(L,x)$, $q = 0, 1, 2, \cdots$, as the starting point of our investigation, and we will denote the corresponding groups by $\tau_q(L,x)$.[7]

It is known that the fundamental group $\pi_1(\mathfrak{V},x)$ arises from the set $M^*(\mathfrak{V},x)$ of all piecewise differentiable loops based at x, with the usual law of multiplication. It follows again from properties (i), (iii) and (iv) that the set $M(\mathfrak{V},x)$ of all standardized loops based at x, with multiplication in P, is equivalent to $M^*(\mathfrak{V},x)$. We can therefore obtain $\pi_1(\mathfrak{V},x)$ from the loop space $M(\mathfrak{V},x)$.

3. Notation and remarks. A little more terminology concerning paths will be helpful. If f, f^* are time-like paths with ff^* defined, we will say that

[5] For basic terminology and facts of differential geometry, in modern form, we refer to K. Nomizu [6].

[6] The inclusion map $I\colon T_q \to T_q^*$ induces a map i of the corresponding groups. Properties (ii) and (iii) imply that i is a homomorphism, (ii) implies that it is an epimorphism, and (iv) implies that it is a monomorphism.

[7] We will generally suppress the symbols (L,x) and write τ_q, etc. whenever the meaning is sufficiently clear.

f joins smoothly to f^*, provided the terminal tangent vector to f and the initial tangent vector to f^* are equally directed. Two paths f, f^* will be called *equal modulo stings* (written $f \doteq f^*$) if one can be obtained from the other by insertion and deletion of stings (applied a finite number of times). An element $f \in T_q(L, x)$ may be called *stingless* if there does not exist a path f^* (different from f) such that f is obtained from f^* by insertion of stings. It is not difficult to see that every $f \in T_q(L, x)$ is equal modulo stings to exactly one stingless element $f^* \in T_q(L, x)$. The operation producing f^* from f may be called *pruning*. Lastly, if f is a path and $t_1, t_2 \in I$, we will denote by $(f \mid t_1, t_2)$ the path g, defined by the formula

$$g(t) = f(t_1 + t(t_2 - t_1)); \qquad t \in I.$$

So far we have employed a Riemannian structure $\mathcal{R}$ only to standardize the parametrization of paths. Apart from this, we shall have occasion to utilize the fact that $\mathcal{R}$ determines a distance function Δ (mapping $\mathfrak{V} \times \mathfrak{V}$ into the set of nonnegative real numbers). For $x \in \mathfrak{V}$ and ϵ a positive number, $\mathfrak{S}(x, \epsilon)$ will denote the totality of points $y \in \mathfrak{V}$ with $\Delta(x, y) < \epsilon$. It will always be understood that ϵ has been chosen sufficiently small to insure that $\mathfrak{S}(x, \epsilon)$ is topologically an open ball (n-cell). In addition to the Riemannian structure $\mathcal{R}$, we shall sometimes utilize a metric Lorentz structure $\mathcal{L}$ associated with L. *Geodesics* and *parallel transport* will be understood relative to $\mathcal{L}$,[8] while *distance* and *path length* will be understood relative to $\mathcal{R}$. Our assumption concerning the differentiability class of $\mathcal{L}$ was designed to insure that, given $x \in \mathfrak{V}$ and a nonzero tangent vector ξ at x, there shall exist a geodesic g which starts at x and has its initial tangent vector in the direction of ξ. Moreover, for ϵ sufficiently small, there exists exactly one such (standardized) g of length ϵ.

With every $f \in M(\mathfrak{V}, x)$ we can associate a Lorentz transformation T_f of the tangent vector space $\mathfrak{V}_x$ at x in the following manner. For $\xi \in \mathfrak{V}_x$, $T_f\xi$ shall be the vector resulting from the parallel transport of ξ around f. If ξ is time-like, so is $T_f\xi$, and one may call T_f *time-preserving* or *time-reversing*, depending on whether ξ and $T_f\xi$ are equally or oppositely directed, respectively. It is not difficult to see that this property of T_f depends only on the homotopy class of f and on the conformal Lorentz structure L associated with $\mathcal{L}$. In the event that T_f is time-preserving for all $f \in M(\mathfrak{V}, x)$, L may be called *time-orientable*. We also define the *parity* of $f \in M(\mathfrak{V}, x)$ to be

[8] More precisely, we should say that there exists a unique torsionless affine connection which preserves $\mathcal{L}$. Geodesics and parallel transport will be understood in terms of this connection.

even or odd, depending on whether T_f is time-preserving or time-reversing. Let $f \in T_q(L, x)$, let $\xi \in \mathfrak{V}_x$ be time-like, and assume that ξ and the initial tangent vector to f are equally directed. Let ξ_t denote the parallel translate of ξ along $(f \mid 0, t)$. If t is a nonsingular point of f, we note that ξ_t and the tangent vector to f at $f(t)$ are equally or oppositely directed, depending on whether the number of interior corners of $(f \mid 0, t)$ is even or odd. The parity of f is consequently equal to the parity of q.

The proofs given in the subsequent sections depend on certain elementary propositions of Lorentz geometry in the small which we now summarize in the form of three Lemmas. The first two refer to the local behavior of geodesics and can be stated as follows:

LEMMA 3.1. *Given $x \in \mathfrak{V}$, there exists a positive number ϵ such that two points of $\mathfrak{S}(x, \epsilon)$ are joined by at most one geodesic lying wholly in $\mathfrak{S}(x, \epsilon)$.*

LEMMA 3.2. *If $x, y \in \mathfrak{V}$ are joined by a unique time-like geodesic, then there exists a positive number ϵ such that every point of $\mathfrak{S}(x, \epsilon)$ is joined to y by a time-like geodesic.*

The third Lemma refers to the deformability of time-like paths. In our terminology it may be formulated as follows:

LEMMA 3.3. *Let f be a 0-loop based at x and let $f \in f(I)$. There exists then an open neighborhood $\mathfrak{N}$ of y such that, for every $z \in \mathfrak{N}$, there exists a 0-loop based at z. If g is an arbitrary path with $f(t) \in g(I)$ for an interior point $t \in I$, then there exists an open neighborhood $\mathfrak{N}$ in $g(I)$ containing $f(t)$ such that f can be deformed as a 0-loop based at x to sweep over $\mathfrak{N}$.*

The proofs of the three propositions offer no difficulty and will be omitted here.

4. Direct limits. The inclusion maps $I_q \colon T_q(L, x) \to M(\mathfrak{V}, x)$, $H_q \colon T_q(L, x) \to T_{q+2}(L, x)$ induce homomorphisms i_q, h_q of the corresponding groups, and we observe that the diagram

$$\begin{tikzcd} \tau_q(L, x) \arrow[rr, "h_q"] \arrow[dr, "i_q"'] & & \tau_{q+2}(L, x) \arrow[dl, "i_{q+2}"] \\ & \pi_1(\mathfrak{V}, x) & \end{tikzcd} \tag{4.1}$$

is commutative. With every base point $x \in \mathfrak{V}$ we have thus associated an algebraic structure which is summarized by diagram (4.1), drawn for

$q = 0, 1, 2, \cdots$. The first point to note is that the maps i_q and h_q are neither monomorphisms nor epimorphisms in general, as will be shown in Section 6 by means of examples. Thus even on a simply-connected manifold the τ-groups may assume nontrivial values. In order to find a connection between the τ-groups and the fundamental group π_1, one must consider the direct limits[9] of

$$\tau_0 \xrightarrow{h_0} \tau_2 \to \cdots \text{ and } \tau_1 \xrightarrow{h_1} \tau_3 \to \cdots.$$

Let us designate these direct limits by τ_+ and τ_-, respectively. Since diagram (4.1) is commutative, the maps i_q determine two homomorphisms $i_\pm : \tau_\pm \to \pi_1$.

THEOREM 4.1. *$i_+(i_-)$ maps $\tau_+(\tau_-)$ isomorphically onto the subgroup $\pi_1^+(\pi_1^-)$ of π_1, generated by loops of even (odd) parity. If L is time-orientable, $\pi_1^+ = \pi_1$ and π_1^-, τ_q (q odd) are trivial. Otherwise π_1 is a normal subgroup of index 2 and $\pi_1^- = \pi_1$.*

We begin the proof by defining sets

$$T_+ = \bigcup_{q \text{ even}} T_q, \qquad T_- = \bigcup_{q \text{ odd}} T_q.$$

We also let $M_+(M_-)$ denote the subset of M generated by loops of even (odd) parity. The purely geometrical content of the theorem may now be summarized in two propositions, as follows:

LEMMA 4.1. *Let $f_0, f_1 \in T_+(T_-)$. If $f_0 \sim f_1$ in M, then there exists an even (odd) integer q such that $f_0 \sim f_1$ in T_q.*

LEMMA 4.2. *Given $f_0 \in M_+(M_-)$, there exists $f_1 \in T_+(T_-)$ such that $f_0 \sim f_1$ in $M_+(M_-)$.*

These assertions are not at all surprising, but they are a little awkward to prove. Let us first consider Lemma 4.1. By assumption there exists a homotopy $F : I \times I \to \mathfrak{V}$ in M with $F_i = f_i$ for $i = 0, 1$. We will show by a constructive argument that there exists an integer q and a homotopy H: $I \times I \to \mathfrak{V}$ in T_q such that $H_i \sim F_i$ in T_q for $i = 0, 1$. The construction will depend on a Riemannian structure $\mathcal{R}$ and a metric Lorentz structure $\mathcal{L}$ associated with L. For $y \in F(I \times I)$ let $S(y)$ denote the set of all positive numbers ϵ such that (i) two points of $\mathfrak{S}(y, 2\epsilon)$ are connected by at most one geodesic lying wholly in $\mathfrak{S}(y, 2\epsilon)$,[10] (ii) $\epsilon \leqq 1$. By Lemma 3.1 $S(y)$ is not

[9] Cf., for instance, Eilenberg and Steenrod [2].

[10] In keeping with an agreement made in Section 3, we are taking ϵ sufficiently small to insure that $\mathfrak{S}(y, 2\epsilon)$ shall be an n-cell.

empty, and we let $\epsilon_1(y)$ denote its least upper bound. Since $\epsilon_1(y)$ is clearly lower semicontinuous, it assumes a smallest (positive) value ϵ_1 on $F(I\times I)$. Now let ξ be a time-like tangent vector at our base point x. With every point $(s,t)\in I\times I$ we associate a time-like tangent vector $\xi_{s,t}$ by parallel translating ξ along $(F_t\,|\,0,s)$. There exists exactly one (standardized) geodesic $G_{s,t}$ which (i) starts at $F(s,t)$, (ii) has initial tangent vector in the direction of $\xi_{s,t}$, (iii) has total length ϵ_1. By Lemma 3.2 there exist positive numbers $\epsilon_{s,t}$, bounded above by ϵ_1, such that $F(s^*,t)\in\mathfrak{S}(F(s,t),\epsilon_{s,t})$ implies the existence of a unique geodesic $H^{s^*}{}_{s,t}$, lying wholly in $\mathfrak{S}(F(s,t),\epsilon_1)$, with

$$H^{s^*}{}_{s,t}(0)=G_{s,t}(1) \text{ and } H^{s^*}{}_{s,t}(1)=F(s^*,t).$$

We observe that the covering $\{\mathfrak{S}(F(s,t),\epsilon_{s,t}),\,(s,t)\in I\times I\}$ of $F(I\times I)$ admits a finite subcovering. There exists consequently a partition of I, given by $0=s_0<s_1<\cdots<s_N=1$, such that

$$F(s_{i+1},t)\in\mathfrak{S}(F(s_i,t),\epsilon_{s_i,t}), \qquad 0\leqq i<N,\ t\in I.$$

We now define the desired homotopy H by the formula

$$H_t=\prod_{i=0}^{N-1}G_{s_i,t}H^{s_{i+1}}{}_{s_i,t}. \tag{4.1}$$

We further define homotopies G^0 and G^1 by

$$G_t{}^i=\prod_{j=0}^{N-1}G_{s_j,i}H^{u_j(t)}{}_{s_j,i}(F_i\,|\,u_j(t),s_{j+1}), \qquad i=0,1, \tag{4.2}$$

where

$$u_j(t)=s_j+t(s_{j+1}-s_j).$$

One observes that

$$G_0{}^i=F_i, \qquad G_1{}^i=H_i, \qquad i=0,1. \tag{4.3}$$

It remains now to prove that G^0, H and G^1 are homotopies in T_q for some q. For every $t\in I$, $G_t{}^0$, H_t and $G_t{}^1$ are time-like paths which start and end at x. We let $q^0(t)$, $p(t)$ and $q^1(t)$, respectively, denote the total number of corners. Since $F_0,F_1\in T_+(T_-)$, there exists an even (odd) integer r such that $F_0,F_1\in T_r$. In virtue of Equation (4.3) we can conclude that $q^0(t)$, $p(t)$ and $q^1(t)$ are even (odd) for all $t\in I$. Moreover, one finds that

$$p(t)\leqq 2N, \qquad q^i(t)\leqq p(t)+r\leqq 2N+r, \qquad i=0,1.$$

Consequently G^0, H and G^1 are homotopies in T_q with

$$q=\max_{t\in I}\{q^0(t),p(t),q^1(t)\},$$

as was to be proved. The proof of Lemma 4. 2 is accomplished by means of the same construction and need not be given here.

We proceed now to prove the theorem. Let $\tau_+^*(\tau_-^*)$ denote the group arising from $T_+(T_-)$. We shall show that $\tau_\pm^*$ is naturally isomorphic to $\tau_\pm$. The latter may be regarded as a set of equivalence classes in $S_\pm$, where

$$S_+ = \bigcup_{q \text{ even}} \tau_q, \qquad S_- = \bigcup_{q \text{ odd}} \tau_q,$$

two elements of $S_\pm$ being called *equivalent* if one can be mapped on the other by a composition of the maps h_q. For $a \in S_\pm$, let $\{a\}$ denote its equivalence class. We define now two maps $\phi_\pm\colon \tau_\pm^* \to \tau_\pm$ by the rule

$$\phi_\pm([l]_\pm) = \{[l]_q\},$$

where l denotes an arbitrary element of $T_\pm$, $[l]_\pm$ denotes the corresponding element in $\tau_\pm^*$, q denotes the smallest integer p such that $l \in T_p$, and $[l]_q$ denotes the corresponding element of τ_q. To show that $\phi_\pm$ is well defined, the corresponding element of τ_q. To show that $\phi_\pm$ is well defined, we suppose $l_0 \sim l_1$ in $T_\pm$. Let q_i denote the smallest integer p such that $l_i \in T_p$, $i = 0, 1$. By Lemma 4. 1 there exists an integer $q \geqq q_0, q_1$ such that $l_0 \sim l_1$ in T_q. Consequently

$$\phi_\pm([l_0]_\pm) = \{[l_0]_{q_0}\} = \{[l_0]_q\} = \{[l_1]_q\} = \{[l_1]_{q_1}\} = \phi_\pm([l_1]_\pm),$$

showing that $\phi_\pm$ is well defined. One verifies by similar elementary considerations that $\phi_\pm$ is an isomorphism, and that the diagram

$$\begin{tikzcd} \tau_\pm \arrow[rr, "i_\pm"] & & \pi_1^\pm \\ & \tau_\pm^* \arrow[ul, "\phi_\pm"] \arrow[ur, "i_\pm^*"'] & \end{tikzcd} \tag{4.4}$$

(where $i_\pm^*$ denotes the homomorphism induced by the inclusion map $I_\pm^*\colon T_\pm \to M_\pm$) is commutative. We claim now that $i_\pm^*$ is an isomorphism. This is so because $i_\pm^*$ must be a monomorphism in virtue of Lemma 4. 1, and it must be an epimorphism by Lemma 4. 2. Since diagram (4. 4) is commutative, we may conclude that $i_\pm$ is likewise an isomorphism.

If L is time-orientable, it is clear that $\pi_1^+ = \pi_1$, and that π_1^-, τ_q (q odd) are trivial. We suppose now that L is not time-orientable. For $l_+ \in M_+$ and $l \in M$, ll_+l^{-1} belongs again to M_+, implying that π_1^+ is normal. It remains to show that π_1^+ has index 2 in π_1, and that $\pi_1^- = \pi_1$. For this purpose we observe that the Riemannian structure $\mathfrak{R}$ permits us to distinguish a special

line element in every tangent vector space $\mathcal{V}_y$ to $\mathcal{V}$ as axis of the light cone. Each axis contains two unit vectors $\xi_y^{\pm}$, one being the negative of the other. The totality of pairs $(\xi_y^{\pm}, y)$, $y \in \mathcal{V}$, may be endowed with a manifold structure, and one thus obtains a manifold $\mathcal{V}^*$ covering $\mathcal{V}$. Since L is not time-orientable, $\mathcal{V}^*$ must be connected. We observe now [11] that $\pi_1(\mathcal{V}^*)$ is isomorphic to $\pi_1^+(\mathcal{V})$, and since $\mathcal{V}^*$ is a 2-fold covering of $\mathcal{V}$, $\pi_1^+(\mathcal{V})$ has index 2 in $\pi_1(\mathcal{V})$. Thus $\pi_1 = \pi_1^+ \cup \pi_1^*$, where π_1^* denotes a coset of π_1^+ in π_1. We note that $\pi_1^* \subset \pi_1^-$, since every element of π_1 which does not belong to π_1^- belongs to π_1^+. Let α be an arbitrary element of π_1 and $\beta \in \pi_1^*$. Then $\alpha\beta^{-1} \in \pi_1^*$, and consequently $\alpha\beta^{-1}$ and β both belong to π_1^-. But this implies $\alpha \in \pi_1^-$, as was to be proved.

5. **Components of L.** In contradistinction to $\pi_1(\mathcal{V}, x)$, the new groups $\tau_q(L, x)$ do in general depend on x as abstract groups.[12] We define a *component* of L to be a maximal connected subset $\mathcal{U}$ of $\mathcal{V}$ such that, for all $x, y \in \mathcal{U}$, there shall exist isomorphisms $\phi_q\colon \tau_q(L, y) \to \tau_q(L, x)$, $q \geqq 0$, which make the following diagram commutative:

$$\begin{array}{ccc} \tau_q(L,y) & \xrightarrow{\phi_q} & \tau_q(L,x) \\ \downarrow{\scriptstyle h_q} & & \downarrow{\scriptstyle h_q} \\ \tau_{q+2}(L,y) & \xrightarrow{\phi_{q+2}} & \tau_{q+2}(L,x). \end{array}$$

This definition implies that the totality of components constitutes a partition (disjoint covering) of $\mathcal{V}$. A Lorentz manifold with just one component may be called *simple*. If x is a point of $\mathcal{V}$, we will denote by $\{x\}$ the component to which it belongs. In this Section we shall establish a geometric characterization of components $\{x\}$ for which $\tau_0(L, x)$ is nontrivial, and we shall show in this case that $\pi_1^+(\{x\}, x)$ is a quotient group of $\tau_0(L, x)$.

THEOREM 5.1. *Let $\tau_0(L, x)$ be nontrivial for some $x \in \mathcal{V}$. Then $\{x\}$ is the totality of points in $\mathcal{V}$ belonging to the range of 0-loops based at x. Moreover, $\tau_0(L, z)$ is trivial for all $z \in \partial\{x\}$*[13].

Let $\mathcal{U}$ denote the totality of points in $\mathcal{V}$ belonging to the range of 0-loops based at x. It follows directly from Lemma 3.3 that every point of $\mathcal{U}$ different from x is an interior point of $\mathcal{U}$. To see that x itself is an interior point,

[11] Cf., also L. Markus [5], p. 412.

[12] Cf., for instance, Example A, Section 6.

[13] For $\mathcal{U} \subset \mathcal{V}$, $\partial\mathcal{U}$ shall denote the boundary of $\mathcal{U}$ in the sense of general topology.

we take a 0-loop f based at x and apply Lemma 3.3 to f^2. Hence $\mathcal{U}$ is open. If $\tau_0(L,z)$ is nontrivial for $z \in \partial\mathcal{U}$, then there exists a 0-loop l based at z. This implies by Lemma 3.3 that z has a neighborhood $\mathcal{N}$ such that, for every $y \in \mathcal{N}$, there exists a 0-loop l^* based at z with $y \in l^*(I)$. Since $z \in \partial\mathcal{U}$, we can take $y \in \mathcal{N} \cap \mathcal{U}$. If $l_+^* l_-^*$ is a y-factorization of l^*, $l_-^* l_+^*$ will be a 0-loop based at y. Since $y \in \mathcal{U}$, there exists a 0-loop f based at x with $y \in f(I)$. Let $f_+ f_-$ be a y-factorization of f. Then either $f_+ l_-^* l_+^* f_-$ or else $f_-^{-1} l_-^* l_+^* f_+^{-1}$ must be a 0-loop based at x, which implies that $z \in \mathcal{U}$. On the other hand, since $\mathcal{U}$ is open, $z \notin \mathcal{U}$. Hence $\tau_0(L,z)$ is trivial. We conclude that $\{x\} \subset \mathcal{U}$.

It remains to show that $\mathcal{U} \subset \{x\}$. Let y be an arbitrary point of $\mathcal{U}$. There exists then a 0-loop l based at x with $y \in l(I)$. Let $l_+ l_-$ be a y-factorization of l, and let q be a nonnegative integer. We wish now to construct a map $\Phi\colon T_q(L,y) \to T_q(L,x)$ which will induce the desired isomorphism ϕ_q of the corresponding groups. To define $\Phi_q(f)$ for an arbitrary element $f \in T_q(L,y)$, we will first prune f to obtain a stingless element f^*. One must now distinguish three cases: (i) $f^* = e_y$, (ii) f^* is a p-loop (based at y) and joins smoothly to l_-, (iii) f^* is a p-loop and joins smoothly to l_+^{-1}. We define $\Phi_q(f)$ in each case as follows:

$$\Phi_q(f) = \begin{cases} \text{(i)} & l_+ f l_+^{-1} \\ \text{(ii)} & l_+ f l_- l^{-1} \\ \text{(iii)} & l l_-^{-1} f l_+^{-1}. \end{cases} \tag{5.1}$$

To prove that $\Phi_q(f) \in T_q(L,x)$, one verifies first that

$$\Phi_q(fg) \doteq \Phi_q(f)\Phi_q(g), \qquad f, g \in T_q(L,y); \tag{5.2}$$

in all cases. It follows by induction on the number ν that

$$\Phi_q\Big(\prod_{i=1}^{\nu} f_i\Big) \doteq \prod_{i=1}^{\nu} \Phi_q(f_i), \qquad f_i \in T_q(L,y). \tag{5.3}$$

Since $f \in T_q(L,y)$, it has a factorization

$$f = \prod_{i=1}^{\nu} f_i, \qquad f_i \in L_q(L,y).^{14} \tag{5.4}$$

In virtue of Equations (5.3) and (5.4), it remains to prove that $\Phi_q(f_i) \in T_q(L,x)$ for $f_i \in L_q(L,y)$. This is certainly true if f_i places us in case (i) as defined above. In case (ii), $l_+ f_i l_-$ is clearly an element of $L_q(L,x)$, and it remains to verify that $l^{-1} \in T_q(L,x)$. To show this, it suffices to observe that $l^{-1} \doteq (l_+ f_i l_-)(l l_+ f_i l_-)^{-1}$, and that $l l_+ f_i l_- \in L_q(L,x)$. Similar observations show that $\Phi_q(f_i) \in T_q(L,x)$ in case (iii), and the map $\Phi_q\colon T_q(L,y) \to T_q(L,x)$ is consequently well defined.

---

[14] $L_q = S(L_q^*)$, as defined in Section 2.

It is not difficult to see that $f \sim g$ in $T_q(L,y)$ implies $\Phi_q(f) \sim \Phi_q(g)$ in $T_q(L,x)$. Therefore Φ_q induces a map $\phi_q\colon \tau_q(L,y) \to \tau_q(L,x)$. Equation (5.2) implies that ϕ_q preserves multiplication, and since $\Phi_q(e_y) = l_+l_+^{-1}$, ϕ_q also preserves the identity. Hence ϕ_q is a homomorphism. To show that it is an isomorphism, we define a map $\Psi_q\colon T_q(L,x) \to T_q(L,y)$ by the formula

$$\Psi_q(f) = \begin{cases} \text{(i)} & l_+^{-1} f l_+ \\ \text{(ii)} & l_+^{-1} l_-^{-1} l_- f l_+ \\ \text{(iii)} & l_+^{-1} f l_-^{-1} l_- l_+, \end{cases} \tag{5.5}$$

where cases (i), (ii) and (iii) are defined just as above, with $l_+(l_-^{-1})$ playing the role of $l_-(l_+^{-1})$. One verifies as before that Ψ_q is well defined, and that it induces a homomorphism $\psi_q\colon \tau_q(L,x) \to \tau_q(L,y)$. A simple calculation based on Equations (5.1) and (5.5) discloses that $\psi_q \circ \phi_q$ and $\phi_q \circ \psi_q$ are identity maps, which proves that ψ_q is an isomorphism.[15]

It remains now to check that the maps ϕ_q commute with the maps h_q. But this follows from the fact that the diagram[16]

$$\begin{array}{ccc} T_q(L,y) & \xrightarrow{\Phi_q} & T_q(L,x) \\ \downarrow & & \downarrow \\ T_{q+2}(L,y) & \xrightarrow{\Phi_{q+2}} & T_{q+2}(L,x) \end{array}$$

is clearly commutative.

Having concluded the proof of Theorem 5.1, we proceed to prove

THEOREM 5.2. *If $\tau_0(L,x)$ is nontrivial, then there exists a natural epimorphism $k\colon \tau_0(L,x) \to \pi_1^+(\{x\},x)$.*

If f is an arbitrary (nontrivial) element of $T_0(L,x)$, we can prune it to obtain an element f^* which is a product of 0-loops based at x. By the preceding theorem, $f^*(I) \subset \{x\}$. We have consequently defined a map $K\colon T_0(L,x) \to M^+(\{x\},x)$, and this induces a homomorphism $k\colon \tau_0(L,x) \to \pi_1^+(\{x\},x)$. Our main task is to verify that an arbitrary element $f \in M^+(\{x\},x)$ is homotopic in $M^+(\{x\},x)$ to a product of 0-loops based at x.

In virtue of Theorem 5.1 there exists, for every $s \in I$, a 0-loop f_s based

[15] To indicate how this calculation goes, we take $f \in T_q(L,y)$ and place ourselves in case (ii). Then

$$\Psi_q \circ \Phi_q(f) = l_+^{-1} l_-^{-1} l_- (l_+ f l_- l_-^{-1} l_+^{-1}) l_+ \doteq f.$$

The five remaining computations are quite similar.

[16] The vertical arrows signify inclusion maps.

at x and an interior point $t_s \in I$ such that $f_s(t_s) = f(s)$. It follows by Lemma 3.3 that every $s \in I$ has an open neighborhood N_s in I, an associated continuous family $\{F_u^s, u \in N_s\}$ of 0-loops based at x, and an associated continuous function u_s: $N_s \to I$ such that $F_t^s(u_s(t)) = f(t)$ for $t \in N_s$, $F_s^s = f_s$ and $u_s(s) = t_s$. One sees by elementary considerations that there exists a finite partition of I, given by $0 = s_0 < s_1 < \cdots < s_\nu = 1$, such that $s_{i-1} \in N_{s_i}$ for $0 < i \leqq \nu$.

We can now simplify our notation by setting $f_i = f_{s_i}$, $F_i = F^{s_i}{}_{s_{i-1}}$, $t_i = t_{s_i}$ and $u_i = u_{s_i}(s_{i-1})$, $0 < i \leqq \nu$. Thus $F_{i+1}(u_{i+1}) = f_i(t_i)$, $0 < i < \nu$, and we may suppose that our loops have been so oriented that $(F_{i+1} \mid 0, u_{i+1})$ does *not* join smoothly to $(f_i \mid t_i, 1)$. We may suppose further that $(F_1 \mid u_1, 1)$ is a 0-loop based at x. Let

$$(5.6) \qquad g_\mu = \prod_{i=1}^{\mu} (F_i \mid u_i, 1)(f_i \mid 1, t_i), \qquad 0 < \mu \leqq \nu.$$

It is not difficult to see, in the first place, that $g_\nu \sim f$ in $M^+(\{x\}, x)$. The formula required to prove this is entirely analogous to Equation (4.2) and need scarcely be written down. One observes further that g_ν is a product involving 0-loops based at x, and possibly some 1-loops based at x with an exterior corner. If the latter do not occur, the result is proved. In the general case we will show that g_ν can be modified by the insertion of stings to yield a product of 0-loops based at x.

For this purpose we posit the inductive hypothesis

$$(5.7) \qquad g_i \doteq l_i m_i (f_i \mid 1, t_i), \qquad 0 < i < \mu \leqq \nu,$$

where l_i denotes a product of 0-loops based at x and m_i is a closed time-like path (without interior corners) which joins smoothly to $(f_i \mid 1, t_i)$, or else $m_i = e_x$. Our assumption concerning $(F_1 \mid u_1, 1)$ insures the validity of this inductive hypothesis. We claim now that a corresponding formula holds for $i = \mu$. This is certainly the case when

$$l = m_{\mu-1}(f_{\mu-1} \mid 1, t_{\mu-1})(F_\mu \mid u_\mu, 1)$$

is a 0-loop, for one can then take $l_\mu = l_{\mu-1} l$ and $m_\mu = e_x$. Otherwise l is 1-loop with exterior corner, and $l_{\mu-1} l \in M^-(\{x\}, x)$. This being the case, we will establish the existence of a closed time-like path m such that lm is a 0-loop based at x. Let us suppose for the moment that this has been done. One can then set $l_\mu = l_{\mu-1} l m$ and $m_\mu = m^{-1}$, and it is easily checked that m_μ joins smoothly to $(f_\mu \mid 1, t_\mu)$, as is required. This completes the inductive argument, and one can conclude that Equation (5.7) holds for $i = \nu$. The

factor $m_\nu(f_\nu \mid 1, t_\nu)$ must be a 0-loop or a 1-loop. Since $l_\nu, g_\nu \in M^+(\{x\}, x)$, the second possibility is ruled out, as was to be proved.

It remains now to construct the path m referred to above. If $\mu = \nu$, then the loop

$$g_\mu{}^* = l_{\mu-1} l(f_\mu \mid 0, t_\mu)$$

must belong to $M^+(\{x\}, x)$, since it differs from g_ν only by a 0-loop and a sting. Consequently $l(f_\mu \mid 0, t_\mu)$ must be a 0-loop, and we can take $m = (f_\mu \mid 0, t_\mu)$. We suppose now $\mu < \nu$. If $\mu + 1 = \nu$, then one observes again that the loop

$$g_{\mu+1}{}^* = l_{\mu-1} l(f_\mu \mid 0, t_\mu)\,(F_{\mu+1} \mid u_{\mu+1}, 0)\,(f_{\mu+1} \mid 1, t_{\mu+1})$$

must belong to $M^+(\{x\}, x)$, since it differs from g_ν only by 0-loops and stings. In this case we may therefore take

$$m = (f_\mu \mid 0, t_\mu)\,(F_{\mu+1} \mid u_{\mu+1}, 0)\,(f_{\mu+1} \mid 1, t_{\mu+1}).$$

If, on the other hand $\mu + 1 < \nu$, then we may continue our construction in the obvious way. After at most $(\nu - \mu)$ steps we will have obtained a suitable path m.

As a corollary to Theorem 5.2, we note that if $\mathfrak{V}$ is simple, then τ_0 is either trivial or it admits $\pi_1{}^+$ as a quotient. In particular, if $\mathfrak{V}$ is simple and compact, then τ_0 admits $\pi_1{}^+$ as a quotient.[17]

6. Examples.

(A) The simplest example of a Lorentz manifold is obtained by endowing the real number plane R^2 with a metric tensor, whose components relative to the standard coordinate system are given by

$$\| g_{ij} \| = \begin{Vmatrix} +1 & 0 \\ 0 & -1 \end{Vmatrix}.$$

Clearly all the τ-groups for this space are trivial. Now let us remove a closed space-like line segment yy^* from R^2 and consider the resulting manifold $\mathfrak{V}$. It is clear that all groups τ_q (q odd) are trivial, because the given Lorentz structure L is time-orientable. It is also evident that τ_0 is everywhere trivial, and one sees without much difficulty that τ_q is everywhere infinite cyclic for $q = 4, 6, 8 \cdots$. This is due to the fact that, given $x \in \mathfrak{V}$, there exists a 4-loop based at x which encircles the missing interval yy^*. The situation is different in regards to τ_2. A given null-line through y will intersect a null-line through y^* at some point z. A second null-line through y will intersect a null-line

[17] We do not presently know whether there exist compact Lorentz manifolds which are not simple.

through y^* at some point z^*. We claim now that $\tau_2(L,x)$ is trivial for all points x belonging to the closed triangular region $y z y^*$ (or $y z^* y^*$). For there does not exist a 2-loop based at x which encircles yy^*. For all other points $x \in \mathfrak{V}$, such a 2-loop does exist, and one finds again that $\tau_2(L,x)$ is infinite cyclic. This example shows that the maps h_q and i_q are not epimorphisms in general.

(B) For our second example we take $\mathfrak{V}$ to be the 3-dimensional number space R^3. A conformal Lorentz structure L may be defined by prescribing a smooth vector field η on R^3. Since R^3 carries a natural Euclidean structure, η determines in every tangent vector space $\mathfrak{V}_x$ a right circular cone with vertex at the origin and axis parallel to $\eta(x)$. We take this to be our light cone. Let r, θ, z be cylindrical coordinates in R^3, and let $(\vec{e}_r, \vec{e}_\theta, \vec{e}_z)$ denote the corresponding local reference frame. We shall set

$$\eta = a_\theta \vec{e}_\theta + a_z \vec{e}_z,$$

where

$$a_z = 1 - a_\theta$$

$$a_\theta = f(r),$$

f being a monotonic C^∞ function such that

$$f(t) = \begin{cases} 0 & t \leqq 0 \\ 1 & t \geqq 1. \end{cases}$$

The resulting Lorentz structure evidently admits 0-loops. In particular, if $x = (1,0,0)$, then the equations

$$r = 1$$

$$\theta = 2\pi t \qquad 0 \leqq t \leqq 1$$

$$z = 0$$

represent a 0-loop based at x. This implies that $\tau_0(L,x)$ is nontrivial. It is extremely likely that the given Lorentz manifold has exactly one component with τ_0 infinite cyclic, all other groups being trivial.

(C) For our last example we take $\mathfrak{V}$ to be the portion of R^3 defined by $|z| < \epsilon$, where ϵ is a positive number. Let x_1, x_2, x_3 denote the natural coordinate numbers in R^3 (with $x_3 = z$) and let $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ denote the corresponding local reference frame. We set

$$\eta = a_2 \vec{e}_2 + a_3 \vec{e}_3$$

with

$$a_3 = 1 - a_2$$

$$a_2 = f(r - 1).$$

Let $x = (3, 0, 0)$, $x' = (0, 4, 0)$, $x'' = (-3, 0, 0)$ and $x''' = (0, -4, 0)$. The The ploygonal line $xx'x''x'''x$ represents now a 2-loop based at x, with corners at x' and x'''. We assert that if ϵ is sufficiently small, then this 2-loop represents a nontrivial element of $\tau_2(L, x)$. For to collapse it, one must make it intersect the line $r = 0$ (the x_3-axis). But it is easily verified that a time-like path which starts and ends at a point with $r > 1$ and intersects the line $r = 0$ cannot have fewer than $(1 + 1/\epsilon)$ corners. This example shows that the maps i_q and h_q are not monomorphisms in general.

A remark concerning *completeness*[18] is in order at this point. The τ-groups and the associated homomorphisms depend only on the conformal Lorentz structure L. Completeness, on the other hand, is a property of the metric Lorentz structure $\mathcal{L}$, and is not a conformal invariant. We believe that completeness by itself is quite irrelevant to the conformal global properties in question. In its stead one requires a purely conformal concept of normality which insures, at least, that the Lorentz manifold is not a fragment of a larger one. For Lorentz structures on the plane, one can define *normality* to mean the absence of certain conformal global objects (called *barriers* in [7]), and one finds (i) two arbitrary normal Lorentz structures are conformally isometric, (ii) every normal Lorentz structure is geodesically connected, (iii) if $\mathcal{L}$ is complete and has an absolutely convergent curvature integral, then the associated conformal structure L is normal. This concept of normality can easily be generalized to arbitrary Lorentz manifolds, and it may turn out to be useful. In particular we do expect that normality, in this sense, will impose some restrictions on the behavior of the τ-groups.

7. Lorentz manifolds which are flat and complete.

In this Section it will be assumed that L has an associated metric Lorentz structure $\mathcal{L}$ which is flat and complete. More precisely, this means that the torsionless affine connection canonically associated with $\mathcal{L}$ is flat and complete.[19] As one may expect, these conditions have a strong implication concerning the τ-groups. The basic result can be stated as follows:

[18] Cf. Section 7.

[19] *Flat* means that the curvature vanishes, and *complete* means that all geodesics are extendible to infinite values of a canonical parameter. Cf. Nomizu [6].

THEOREM 7.1. *If $\mathfrak{V}$ is flat and complete,*[20] *then the maps h_0 are monomorphisms and the maps h_q, $q > 0$, are isomorphisms.*

Let $\mathfrak{V}^*$ denote the universal covering space of $\mathfrak{V}$, endowed with the metric Lorentz structure induced by $\mathfrak{V}$. Let $\mathfrak{L}^n$ denote what may be called *standard Lorentz space* of dimension n. It is obtained by endowing the number space R^n with a metric tensor, whose components relative to the standard coordinate system are given by

$$\| g_{ij} \| = \begin{Vmatrix} 1 & 0 \cdots & 0 \\ 0 & -1 \cdots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 \cdots & -1 \end{Vmatrix}.$$

Since $\mathfrak{V}^*$ is flat, complete and simply connected, there exists[21] a global isometry $\phi\colon \mathfrak{V}^* \to \mathfrak{L}^n$. We will henceforth identify $\mathfrak{V}^*$ with $\mathfrak{L}^n$.

To prove that $h_q\colon \tau_q(\mathfrak{V}, x) \to \tau_{q+2}(\mathfrak{V}, x)$ is a monomorphism for $q \geqq 0$ and $x \in \mathfrak{V}$, we must show that $f \nsim e_x$ in $T_q(\mathfrak{V}, x)$ implies $f \nsim e_x$ in $T_{q+2}(\mathfrak{V}, x)$, for $f \in T_q(\mathfrak{V}, x)$. It suffices to show that $f \sim e_x$ in $M(\mathfrak{V}, x)$ implies $f \sim e_x$ in $T_q(\mathfrak{V}, x)$. Let us therefore suppose that f is null-homotopic. If $y_0 \in \mathfrak{V}^*$ lies above x, then f lifts to an element $f^* \in T_q(\mathfrak{V}^*, y_0)$. Now let $H_t\colon \mathfrak{V}^* \to \mathfrak{V}^*$ denote the homothety map defined by the formula[22]

$$H_t(y) = y_0 + t(y - y_0) \tag{7.1}$$

for $t \in I$ and $y \in \mathfrak{V}^*$. Clearly $H_t \circ f^*$ defines a homotopy in $T_q(\mathfrak{V}^*, y_0)$, connecting f^* and e_{y_0}. Let π denote the projection from $\mathfrak{V}^*$ to $\mathfrak{V}$. Then $\pi \circ H_t \circ f^*$ defines a homotopy in $T_q(\mathfrak{V}, x)$, connecting f and e_x.

It remains to show that the maps h_q are epimorphisms for $q > 0$. Given $f \in T_{q+2}(\mathfrak{V}, x)$, we must show that there exists $g \in T_q(\mathfrak{V}, x)$ such that $f \sim g$ in $T_{q+2}(\mathfrak{V}, x)$. It suffices to consider the case where f is a $(q+2)$-loop. If $f \sim e_x$ in $T_{q+2}(\mathfrak{V}, x)$, the result is trivial. We suppose therefore that $f \nsim e_x$ in $T_{q+2}(\mathfrak{V}, x)$, which implies (by the result just proved) that f is not null-homotopic. The lift f^* of f, determined by the base point y_0, will consequently end at a point y^* different from y_0. Clearly there exists a polygonal line of time-like segments $l^* = y_0 y_1 \cdots y_\nu y^*$ in $\mathfrak{V}^*$, with noncollinear vertices, such

[20] In this Section it will be convenient to let $\mathfrak{V}$ denote the differentiable manifold, together with $\mathfrak{L}$.

[21] Cf. Ehresmann [1].

[22] To interpret Equation (7.1), we recall that $\mathfrak{V}^*$ is identified with $\mathfrak{L}^n$, which has a natural vector space structure.

that $f \sim \pi \circ l^*$ in $T_{q+2}(\mathfrak{V}, x)$. One may further assume that the points y_0, $y_{\nu-1}$, y_ν and y^* are coplanar. We take ν to be minimal and assert that $\nu = 0$ or 1. For let L denote the line through y_ν and y^*, and suppose $\nu > 0$. The segment $H_t \circ (y_{\nu-1} y_\nu)$, $0 < t \leqq 1$, can be extended to meet L at a point z_t. As t varies from 1 to 0, z_t varies from y_ν to a limiting position z_0. Clearly y^* cannot lie between y_ν and z_0 on L, for this would contradict the minimality of ν. Consequently

$$\pi \circ \{[H_t \circ (y_0 y_1 \cdot \cdot \cdot y_\nu)][H_t(y_\nu) z_t y^*]\}$$

defines a homotopy in $T_{q+2}(\mathfrak{V}, x)$ connecting $\pi \circ l^*$ with $\pi \circ (y_0 z_0 y^*)$. Thus if $\nu > 0$, one may conclude that $\nu = 1$. $\pi \circ l^*$ is therefore a p-loop, the possible values of p being 0, 1, and 2. If q is odd, $p = 1$ and $\pi \circ l^* \in T_q(\mathfrak{V}, x)$. Otherwise $p = 0$ or 2, and again $\pi \circ l^* \in T_q(\mathfrak{V}, x)$. This proves that the maps h_q are epimorphisms for $q > 0$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] Ch. Ehresmann, "Les connexions infinitesimales dans un espace fibré différentiable," *Colloque de Topologie* (Espaces Fibrés), Bruxelles, 1950.

[2] S. Eilenberg and N. Steenrod, *Foundations of algebraic topology*, Princeton University Press, 1952.

[3] D. Finkelstein and C. W. Misner, "Some new conservation laws," *Annals of Physics*, vol. 6 (1959), pp. 230-243.

[4] P. J. Hilton, *An introduction to homotopy theory*, Cambridge University Press, 1953.

[5] L. Markus, "Line element fields and Lorentz structures on differentiable manifolds," *Annals of Mathematics*, vol. 62 (1955), pp. 411-417.

[6] K. Nomizu, *Lie groups and differential geometry*, Mathematical Society of Japan, 1956.

[7] J. W. Smith, "Lorentz structures on the plane," *Transactions of the American Mathematical Society*, vol. 95 (1960), pp. 226-237.

[8] ———, "Fundamental groups on a Lorentz manifold," *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 46 (1960), pp. 111-114.

ANALYTIC AND ALGEBRAIC DEPENDENCE OF MEROMORPHIC FUNCTIONS.*

By REINHOLD REMMERT.†

1. Let V be an irreducible, locally irreducible complex space[1] and $m_1,\cdots,m_k, m$ meromorphic functions on V such that m is analytically dependent on $m_1,\cdots,m_k$. It is a classical problem to find sufficient conditions under which such a dependence is an algebraic one. A well-known theorem states that analytic dependence always entails algebraic dependence if V is compact (compare [5], [13]). In his papers [10], [12] W. Thimm pointed out another sufficient condition expressed in terms of the structure of the set of indeterminacy of the functions $m_1,\cdots,m_k$. Here we prove a theorem which covers the results of [5], [13] as well as those of [10], [12].

In order to make things clearer we first recall some basic notions. If m_1 is a meromorphic function on V, then the set A_1 of zeros of m_1 and the set B_1 of poles of m_1 is well defined, and both are analytic sets in V which are empty or of codimension 1 everywhere.[2] The intersection $N_1 := A_1 \cap B_1$ is called the set of indeterminacy of m_1, N_1 is an analytic subset of V of codimension at least 2. If $\dim V > 1$, in general N_1 is not empty. For a finite system $m_1,\cdots,m_k$ of meromorphic functions on V we denote by N_κ the set of indeterminacy of m_κ, $\kappa = 1,\cdots,k$, and call $N := \bigcup_{\kappa=1}^{k} N_\kappa$ the set of indeterminacy of the system $m_1,\cdots,m_k$; obviously N is an analytic set in V of codimension at least 2. We denote by S_k the complex product manifold of k complex projective lines and consider the holomorphic map $m_1 \times \cdots \times m_k \colon V - N \to S_k$ defined by $v \to (m_1(v),\cdots,m_k(v))$, $v \in V - N$. This map cannot be continued to a holomorphic map of the total space V

* Received April 5, 1960.

† Part of this paper was developed during the summer 1959 while the author was at the University of Michigan, Ann Arbor.

[1] For the definition of complex spaces compare [2].

[2] By dimension and codimension of complex spaces (resp. analytic sets) we always mean complex dimension and codimension. We denote by $\dim_x X$ the (complex) dimension of a complex space X at a point $x \in X$ and write $\dim X = \max_{x \in X} \dim_x X$. The topological dimension of a topological space X will be denoted by $\dim_R X$.

into S_k, nevertheless for any point $v \in N$ the image set $m_1 \times \cdots \times m_k(v)$ is well defined: it consists of all points $s \in S_k$ which are cluster points of sequences $m_1 \times \cdots \times m_k(v_n)$, where $v_n \in V - N$ tends to v. In this way the meromorphic functions $m_1, \cdots, m_k$ give rise to the correspondence

$$v \to m_1 \times \cdots \times m_k(v) \subset S_k, v \in V,$$

which is called the meromorphic map defined by $m_1, \cdots, m_k$. For any set $M \subset V$ we define the image $m_1 \times \cdots \times m_k(M)$ to be the set

$$\bigcup_{v \in M} m_1 \times \cdots \times m_k(v).$$

For convenience we repeat the definitions of analytic and algebraic dependence. A meromorphic function m on V is called *analytically dependent* on the meromorphic functions $m_1, \cdots, m_k$ if there exist a nonempty domain $U \subset V$ and a holomorphic function $f(w, z_1, \cdots, z_k)$ in a domain G of the space C^{k+1} of the variables $w, z_1, \cdots, z_k$ such that the following is true: $f'_w(w, z_1, \cdots, z_k) \not\equiv 0$; the restrictions $m \mid U, m_1 \mid U, \cdots, m_k \mid U$ are holomorphic in U; $m \times m_1 \times \cdots \times m_k(U) \subset G$; $f(m(v), m_1(v), \cdots, m_k(v)) = 0$ for each point $v \in U$.

m is called *algebraically dependent* on $m_1, \cdots, m_k$ if there exists a polynomial $p(w, z_1, \cdots, z_k)$ with $p'_w(w, z_1, \cdots, z_k) \not\equiv 0$ such that $p(m(v), m_1(v), \cdots, m_k(v)) = 0$ for each point $v \in V$ at which all the functions $m, m_1, \cdots, m_k$ are holomorphic.

If K is any set in V, the notion of a germ of a meromorphic function in K is well defined. Let us denote the set of all germs of meromorphic functions in K by R_K. Obviously R_K forms a ring with respect to the natural addition and multiplication of germs. In the following we always assume K to be connected, in which case R_K is a field. The notions of analytic and algebraic dependence carry over in a natural way to germs of meromorphic functions.

If $m_1, \cdots, m_k$ are meromorphic functions in an open set U of V, we denote their corresponding germs on a given connected set $K \subset U$ by $\bar{m}_1, \cdots, \bar{m}_k$. These germs generate a subfield of R_K which is usually denoted by $C(\bar{m}_1, \cdots, \bar{m}_k)$. The set L_K of all germs of meromorphic functions on K which depend analytically on $\bar{m}_1, \cdots, \bar{m}_k$ obviously forms a subfield of R_K containing $C(\bar{m}_1, \cdots, \bar{m}_k)$; we call L_K the *analytic closure* of $C(\bar{m}_1, \cdots, \bar{m}_k)$ in R_K. We are interested in the algebraic structure of this field L_K. In particular, we want to derive a condition insuring that L_K is an algebraic extension of the field $C(\bar{m}_1, \cdots, \bar{m}_k)$. We prove the following

THEOREM. *Let K be a connected compact analytic set in an irreducible,*

locally irreducible complex space V with a countable base of open sets, and let $m_1, \cdots, m_k$ be meromorphic functions in a neighborhood U of K such that $\dim_R(m_1 \times \cdots \times m_k(U)) = \dim_R(m_1 \times \cdots \times m_k(K))$.[2] *Let us denote by $\bar{m}_1, \cdots, \bar{m}_k$ the germs of meromorphic functions in K defined by $m_1, \cdots, m_k$ on K. Then the analytic closure L_K of $C(\bar{m}_1, \cdots, \bar{m}_k)$ in R_K coincides with the algebraic closure of $C(\bar{m}_1, \cdots, \bar{m}_k)$ in R_K and is a simple algebraic extension of $C(\bar{m}_1, \cdots, \bar{m}_k)$.*[3]

Let us explicitly state the following

COROLLARY. *Let $m_1, \cdots, m_k$ be meromorphic functions on an irreducible, locally irreducible*[3] *complex space V; let $K \subset V$ be a compact analytic set such that* $\dim_R(m_1 \times \cdots \times m_k(K)) = \dim_R(m_1 \times \cdots \times m_k(V))$. *Then the field L of all meromorphic functions on V which depend analytically on $m_1, \cdots, m_k$ is a simple algebraic extension of the field $C(m_1, \cdots, m_k)$.*

In particular this corollary tells us that under the condition $m_1 \times \cdots \times m_k(K) = m_1 \times \cdots \times m_k(V)$ analytic dependence on $m_1, \cdots, m_k$ always entails algebraic dependence on $m_1, \cdots, m_k$. If V is compact then the condition of the corollary is fulfilled for $K = V$ and we have the results of [5], [13]. For the case when K is a single point $v_0 \in V$ the above theorem is essentially due to W. Thimm (compare [12], p. 156, verallgemeinerter Normalfall), who gave a rather complex proof. The following very special case of the corollary was also proved by K. Stein [8]:[4]

If m_1 is a meromorphic function on an irreducible, locally irreducible[3] *complex space V having at least one point of indeterminacy (i.e. N_1 is not empty), then any meromorphic function on V, which depends analytically on m_1, depends algebraically on m_1.*

In the following sections we give a simple proof of the above theorem (and its corollary) by using some results concerning holomorphic maps. The crucial points of the proof are a general lemma on holomorphically dependent holomorphic maps, which is proved in Section 2, and a well-known theorem of W. L. Chow [1] stating that any analytic set in S_k is algebraic.[5]

2. By X, Y, Z we always denote finite dimensional complex spaces with

[3] The condition that V is locally irreducible is only for convenience. If V is not locally irreducible one can always pass to the normalisation of V.

[4] Recently I was informed by K. Stein that he was able to prove the general theorem of Thimm by applying general results of his theory of complex base spaces (unpublished).

[5] For a simple proof of Chow's theorem see [6].

countable bases of open sets. Let $\eta: X \to Y$ be a holomorphic map. The rank $r_x(\eta)$ of η at a point $x \in X$ is defined to be the integer $\dim_x X - \dim_x \eta^{-1}(\eta(x))$, i. e. the codimension at x of the fibre $\eta^{-1}(\eta(x))$. The number $r(\eta) := \max_{x \in X} r_x(\eta)$ is called the (total) *rank* of η on X. Obviously we have $r(\eta) \leqq \dim X$. In [4], [7] it was proved that:

(a) *The image $\eta(X)$ is a topological space of topological dimension $2r(\eta)$. If X is compact, $\eta(X)$ is an analytic set in Y.*

From now on X is supposed to be irreducible. Besides $\eta: X \to Y$ we consider a second holomorphic map $\zeta: X \to Z$. We denote the map $x \to (\eta(x), \zeta(x))$ by $\eta \times \zeta$; obviously $\eta \times \zeta: X \to Y \times Z$ is also a holomorphic map. As $(\eta \times \zeta)^{-1}((\eta \times \zeta)(x)) = \eta^{-1}(\eta(x)) \cap \zeta^{-1}(\zeta(x))$ for all $x \in X$, we have $r_x(\eta) \leqq r_x(\eta \times \zeta)$ for all $x \in X$ and therefore especially $r(\eta) \leqq r(\eta \times \zeta)$. *The map ζ is called holomorphically dependent*[6] *on η if $r(\eta) = r(\eta \times \zeta)$.* This notion of holomorphic dependence of holomorphic maps generalises the well-known notion of analytic dependence of holomorphic functions. For it can easily be proved that:

(b) *A holomorphic function f on X depends analytically on a system $f_1, \cdots, f_k$ of holomorphic functions on X if and only if the equation $r(f_1 \times \cdots \times f_k) = r(f_1 \times \cdots \times f_k \times f)$* holds for the holomorphic maps $f_1 \times \cdots \times f_k: X \to C^k$, $f: X \to C^1$.

Now we prove the following

LEMMA. *Let $\eta: X \to Y$, $\zeta: X \to Z$ be holomorphic maps of an irreducible complex space X into complex spaces Y, Z; let W be a subset of X such that $\dim_R(\eta(W)) = \dim_R(\eta(X))$. Suppose ζ depends holomorphically on η and $\eta \times \zeta(W)$ is an analytic set in $Y \times Z$. Then $\eta \times \zeta(X) = \eta \times \zeta(W)$.*

Proof. We have to prove $\tau^{-1}(\tau(W)) = X$, if we put $\tau := \eta \times \zeta$. Evidently $X' := \tau^{-1}(\tau(W))$ is an analytic set in X containing W. X being irreducible, it is sufficient to prove $\dim X' = \dim X$. Suppose $\dim X' < \dim X$. We consider the restriction maps $\eta' := \eta \mid X'$, $\zeta' := \zeta \mid X'$, $\tau' := \tau \mid X'$. As $\tau' = \eta' \times \zeta'$ and all fibres of τ' are also fibres of τ we get:

$$r_x(\eta') \leqq r_x(\tau') = \dim_x X' - \dim_x \tau'^{-1}(\tau'(x)) < \dim_x X - \dim_x \tau^{-1}(\tau(x)) = r_x(\tau)$$

for all $x \in X'$ and hence $r(\eta') < r(\tau)$. Now $r(\tau) = r(\eta)$ because ζ depends holomorphically on η. Therefore it follows $r(\eta') < r(\eta)$ and by using (a):

[6] This definition is due to K. Stein [9].

$\dim_R(\eta'(X')) < 2r(\eta)$. But this is impossible in view of $\eta'(X') = \eta(X') \supset \eta(W)$ and $\dim_R(\eta(W)) = 2r(\eta)$, q. e. d.

3. For the proof of our theorem we shall use the fact that the points of indeterminacy of a system of meromorphic functions can be removed by a modification. More exactly we shall need the following proposition:

(c) *For each system $m_1, \cdots, m_k$ of meromorphic functions on an irreducible, locally irreducible complex space V having the set N as set of indeterminacy there exists an irreducible, locally irreducible complex space V' and a proper surjective holomorphic map $\nu\colon V' \to V$ with the following properties:*

α) *The topological closure $\bar{V}$ of the graph of the holomorphic map $m_1 \times \cdots \times m_k \mid V - N \to S_k$ in $V \times S_k$ is an irreducible analytic set in $V \times S_k$. V' is the normalisation space of $\bar{V}$, ν is the natural projection of V' onto V.*

β) *ν maps $V' - \nu^{-1}(N)$ biholomorphically onto $V - N$ and induces an isomorphism $\nu^*\colon k(V) \to k(V')$ of the field $k(V)$ of meromorphic functions on V onto the field $k(V')$ of meromorphic functions on V'. ($\nu^*(m) := m \circ \nu$ for each $m \in k(V)$.)*

γ) *The functions $\nu^*(m_1), \cdots, \nu^*(m_k)$ have no points of indeterminacy in V'.*

δ) *For each system $g_1, \cdots, g_r \in k(V)$ and each set $M \subset V$ we have*

$$g_1 \times \cdots \times g_r(M) = \nu^*(g_1) \times \cdots \times \nu^*(g_r)(\nu^{-1}(M)).$$

The statements α)-γ) are proved in [4]; the proof of δ) runs as follows: We write for abbreviation $g := g_1 \times \cdots \times g_r$, $g' := \nu^*(g_1) \times \cdots \times \nu^*(g_r)$ and show first that $g(M) \subset g'(\nu^{-1}(M))$. Suppose $s \in g(v)$, $v \in M$. If we denote by N_g the set of indeterminacy of the functions $g_1, \cdots, g_r$, then by definition there exists a sequence $v_n \in V - N_g$ tending to v such that s is a cluster point of the sequence $g(v_n)$. For each n we choose a point $v'_n \in \nu^{-1}(v_n)$; as ν is a proper map, the sequence $v'_n \in V' - \nu^{-1}(N_g)$ has at least one cluster point v' in V'. Obviously $v' \in \nu^{-1}(v)$. Now we have $g' = g \circ \nu$ on $V' - \nu^{-1}(N_g)$, and therefore $g'(v'_n) = g(v_n)$. This shows that s is a cluster point of $g'(v'_n)$; i. e. $s \in g'(v')$ or $s \in g'(\nu^{-1}(M))$.

The proof of $g'(\nu^{-1}(M)) \subset g(M)$ follows in the same way.

A simple consequence of (a) and (c) is

(c′) *For each compact analytic set K in V the set $m_1 \times \cdots \times m_k(K) \subset S_k$ is algebraic.*

Proof. The map $m' \colon V' \to S_k$ is holomorphic ($m' := \nu^*(m_1) \times \cdots \times \nu^*(m_k)$). Therefore, $\nu^{-1}(K)$ being compact, $m'(\nu^{-1}(K))$ is analytic in S_k by virtue of (a). From (c) and Chow's theorem we then conclude that $m'(\nu^{-1}(K)) = m_1 \times \cdots \times m_k(K)$ is algebraic in S_k.[7]

4. In this section we prove first our theorem and then its corollary. Obviously we may assume $V = U$ without loss of generality. First we show that the field L_K is the algebraic closure of $C(\bar{m}_1, \cdots, \bar{m}_k)$ in R_K. To prove this we have only to show that each $\bar{m} \in L_K$ is algebraic over $C(\bar{m}_1, \cdots, \bar{m}_k)$. We fix an element $\bar{m} \in L_K$ and choose a meromorphic function m in a suitable connected neighbrohood T of K which induces $\bar{m}$ on K and depends analytically on $m_1 \mid T, \cdots, m_k \mid T$. Then we are through if we show that m depends algebraically on $m_1 \mid T, \cdots, m_k \mid T$. From (c′) it follows easily that this statement is contained in the following

PROPOSITION. *Let $m_1, \cdots, m_k$ be a system of meromorphic functions on an irreducible, locally irreducible*[8] *complex space V and let K be a compact analytic set in V such that*

$$\dim_R(m_1 \times \cdots \times m_k(K)) = \dim_R(m_1 \times \cdots \times m_k(V)).$$

Then for any meromorphic function m in a connected neighborhood T of K depending analytically on $m_1 \mid T, \cdots, m_k \mid T$ we have:

$$m_1 \times \cdots \times m_k \times m(T) = m_1 \times \cdots \times m_k \times m(K);$$
$$\dim_R(m_1 \times \cdots \times m_k \times m(T)) = \dim_R(m_1 \times \cdots \times m_k(T)).$$

Proof. We apply (c) to the system $m_1 \mid T, \cdots, m_k \mid T, m$ defined on the irreducible and locally irreducible complex space T. We denote the modification by T' and the modification map by $\nu \colon T' \to T$ and write for abbreviation $\mu' := \nu^*(m_1 \mid T) \times \cdots \times \nu^*(m_k \mid T)$, $m' := \nu^*(m)$. The maps $\mu' \colon T' \to S_k$ and $m' \colon T' \to S_1$ are both holomorphic, and by assumption m' depends holomorphically on μ'. The set $K' := \nu^{-1}(K)$ is compact and analytic in T'; therefore by virtue of (c′), $\mu' \times m'(K')$ is an analytic set in S_{k+1}. As moreover we have $\dim_R(\mu'(K')) = \dim_R(\mu'(T'))$, we are able to apply the lemma of Section 2 (put $X := T'$, $Y := S_k$, $Z := S_1$, $W := K'$, $\eta := \mu'$, $\zeta := m'$). We conclude that $\mu' \times m'(K') = \mu' \times m'(T')$, or what is the same:

---

[7] This was already proved in [11], if K is a single point.

$$m_1 \times \cdots \times m_k \times m(T) = m_1 \times \cdots \times m_k \times m(K).$$

As

$$\dim_R(m_1 \times \cdots \times m_k \times m(T)) = 2r(\mu' \times m'),$$
$$\dim_R(m_1 \times \cdots \times m_k(T)) = 2r(\mu') \text{ and } r(\mu' \times m') = r(\mu'),$$

the second equation of the proposition follows immediately. Q. E. D.

It remains to prove that the field L_K is a simple extension of $C(\bar{m}_1, \cdots, \bar{m}_k)$. A theorem of classical algebra tells us that to prove this it is enough to show that there exists an integer $b > 0$ such that

$$(*) \qquad [\bar{m}: C(\bar{m}_1, \cdots, \bar{m}_k)] \leqq b \text{ for all } \bar{m} \in L_K.$$

We may assume without loss of generality that the set of indeterminacy of the system $m_1, \cdots, m_k$ is empty.[8]. For otherwise we just apply (c) and prove our statement for the fields

$$L_{K'} := \nu^*(L_K),\ C(\bar{m}'_1, \cdots, \bar{m}'_k) := \nu^*(C(\bar{m}_1, \cdots, \bar{m}_k)),$$

which are isomorphic to L_K and $C(\bar{m}_1, \cdots, \bar{m}_k)$ respectively. We denote the holomorphic map $m_1 \times \cdots \times m_k$ of V into S_k by μ and consider the set $B := \mu(V) \subset S_k$. The lemma of Section 2—applied to the map $\eta := \mu$ and a constant map ζ (with $W = K$)—immediately gives $\mu(V) = \mu(K)$; therefore the set B is an irreducible complex subspace of S_k. Obviously no set $\mu^{-1}(s) \cap K$, $s \in B$, is empty. As K is compact, there are only a finite number of components of each fibre $\mu^{-1}(s)$, $s \in B$, which intersect K. Let us denote this finite number by $n(s)$. A well-known proposition of Osgood ([3], p. 230) states the existence of an integer $b > 0$ and of a nonempty domain $G \subset B$ such that the set $G_1 := \{s \in G, n(s) \leqq b\}$ is dense in G. We claim that b is a number fulfilling the conditions of inequality (*).

For the proof let $\bar{m} \in L_K$ be arbitrary. We choose a connected neighborhood T of K and a representative meromorphic function m of $\bar{m}$ in T such that the conditions of our proposition are fulfilled. Then the set $A := (\mu \mid T \times m)(T) \subset S_k \times S_1$ is an irreducible complex subspace of $S_k \times S_1$. The restriction of the natural projection $S_k \times S_1 \to S_k$ to A defines a holomorphic surjective map $\pi: A \to B$. From the equation $\dim A = \dim B$ it follows that there exists a set N analytic and nowhere dense in B such that all sets $\pi^{-1}(s)$, $s \in B - N$, consist of a finite number of points.[9] This shows

[8] The following proof runs along almost the same lines as the proof of Satz 7 in the author's paper [5].

[9] Obviously N can be chosen as the π-image of the degeneration set of the map $\pi: A \to B$.

that the space $A - \pi^{-1}(N)$ is an "analytic covering space" of $B - N$ with respect to the map π.[10] Therefore the inequality $[\bar{m}: C(\bar{m}_1, \cdots, \bar{m}_k)] \leqq b$ is equivalent to the fact that the number of sheets of this covering does not exceed b.

In order to prove this we denote the set of indeterminacy of m in T by M. All sets $\mu^{-1}(s) \cap M$ are empty for $s \in B - N$, as otherwise for a certain $s \in B - N$ the set $\pi^{-1}(s) = (\mu \mid T \times m)(\mu^{-1}(s) \cap T)$, which coincides with $(\mu \mid T)(\mu^{-1}(s) \cap T) \times m(\mu^{-1}(s) \cap T)$, would not be finite.[11] From the algebraic dependence of m on $m_1 \mid T, \cdots, m_k \mid T$ we now conclude that m is constant on each connected component of $\mu^{-1}(s) \cap T$, $s \in B - N$. Therefore it follows from $\mu \times m(K) = \mu \times m(T)$ that m has at most $n(s)$ different values, including ∞, on $\mu^{-1}(s) \cap T$, if $s \in B - N$. Therefore m has always at most b different values on the fibres $\mu^{-1}(s) \cap T$, $s \in G' - N$. But as $G' - N$ is everywhere dense in G, it follows immediately that for each s of the domain $G \subset B$ the set $\pi^{-1}(s)$ consists of at most b different points. This means exactly that the complex space $A - \pi^{-1}(N)$, considered as a covering space of $B - N$ with respect to π, has at most b sheets. Therefore the inequality (*) and hence the theorem is proved.

The proof of the corollary is now very simple. We may assume K to be irreducible. For the analytic set $m_1 \times \cdots \times m_k(V)$, being the holomorphic image of an irreducible complex space, is irreducible, and so there exists an irreducible component K_1 of K such that

$$m_1 \times \cdots \times m_k(K_1) = m_1 \times \cdots \times m_k(V).$$

We then have a natural (injective) field homomorphism of L into L_K which maps $C(m_1, \cdots, m_k)$ onto $C(\bar{m}_1, \cdots, \bar{m}_k)$. The field L therefore may be considered as a subfield of L_K containing $C(\bar{m}_1, \cdots, \bar{m}_k)$. As L_K is a simple algebraic extension of $C(\bar{m}_1, \cdots, \bar{m}_k)$ by our theorem, L is itself a simple algebraic extension of the field $C(m_1, \cdots, m_k)$, Q. E. D.

THE UNIVERSITY OF MICHIGAN
AND
INSTITUTE FOR ADVANCED STUDY.

---

[10] More precisely, the normalisation of $A - \pi^{-1}(N)$ is an analytic covering space of the normalisation of $B - N$. For the notion of analytic covering space compare [2].

[11] The identity $(\mu \mid T \times m)(D) = (\mu \mid T)(D) \times m(D)$ holds for each subset D of T, as $\mu \mid T$ is a holomorphic map.

REFERENCES.

[1] W. L. Chow, "On compact analytic varieties," *American Journal of Mathematics*, vol. 71 (1949), pp. 893-914.

[2] H. Grauert and R. Remmert, "Komplexe Räume," *Mathematische Annalen*, vol. 136 (1958), pp. 245-318.

[3] W. F. Osgood, *Lehrbuch der Funktionentheorie II, 1*, Berlin u. Leipzig, 1929.

[4] R. Remmert, "Holomorphe und meromorphe Abbildungen komplexer Räume," *Mathematische Annalen*, vol. 133 (1957), pp. 328-370.

[5] ———, "Meromorphe Funktionen in kompakten komplexen Räumen," *ibid.*, vol. 132 (1956), pp. 277-288.

[6] ——— and K. Stein, "Über die wesentlichen Singularitäten analytischer Mengen," *ibid.*, vol. 126 (1953), pp. 263-306.

[7] ———, "Eigentliche holomorphe Abbildungen," *Mathematische Zeitschrift*, vol. 73 (1960), pp. 159-189.

[8] K. Stein, "Analytische Projektion komplexer Mannigfaltigkeiten," *Coll. Fonct. Plus. Var.*, Brussels, 1953.

[9] ———, "Analytische Zerlegungen komplexer Räume," *Mathematische Annalen*, vol. 132 (1956), pp. 63-93.

[10] W. Thimm, "Über ausgeartete meromorphe Abbildungen II," *ibid.*, vol. 125 (1952), pp. 264-283.

[11] ———, "Über die Menge der singulären Bildpunkte einer meromorphen Abbildung," *Mathematische Zeitschrift*, vol. 57 (1953), pp. 456-480.

[12] ———, "Untersuchungen über ausgeartete meromorphe Abbildungen," *Mathematische Annalen*, vol. 127 (1954), pp. 150-161.

[13] ———, "Meromorphe Abbildungen von Riemannschen Bereichen," *Mathematische Zeitschrift*, vol. 60 (1954), pp. 435-457.

HEREDITARILY COMPACT SPACES.*

By A. H. Stone.

1. Introduction. By definition, a *hereditarily compact* space, or a "Zariski" or "Noether" space, is a topological space all of whose subspaces are compact.[1] Such spaces have received some attention [9, 10] because they arise in algebraic geometry (in the Zariski topology) and in some other algebraic constructions. Here we study these spaces on their own account. In the applications they are usually T_1 but not T_2; in fact, a T_2 hereditarily compact space is necessarily finite. However, we do not assume any separation axioms except where they are explicitly stated. We begin by giving some alternative characterizations (§ 2), and considering some properties related to some of them (§ 3). In § 4 we associate to every hereditarily compact space a topologically invariant ordinal number, its "type"; this corresponds to the dimension in the application to algebraic geometry. This permits the "construction" of all hereditarily compact spaces (§ 5). In § 6 we discuss the effect of various standard operations on such spaces on their types, and in § 7 we consider the countable hereditarily compact spaces in more detail.

Notation. A space X is *discrete* if each point $p \in X$ has a neighborhood in X consisting of p itself; X is *weakly discrete* if each $p \in X$ has a neighborhood in X consisting of a finite set of points. (For T_1 spaces these notions are equivalent.) An indexed family $\{U_\lambda\}$ of subsets of X is called *finite* if $U_\lambda = \emptyset$ for all but finitely many values of λ.

2. Characterizations. We begin by observing that, in the definition of hereditary compactness, it is not necessary to specify that *all* subspaces are compact, and moreover the kind of compactness considered does not greatly matter. Incidentally we also obtain some further characterizations.

THEOREM 1. *The following statements about an arbitrary space X are equivalent:*

* Received April 25, 1960.

[1] Throughout this paper, "compact" means "quasicompact" in the sense of Bourbaki; that is, every open covering has a finite subcovering.

(1) *Every subspace of X is compact.*

(1_c) *Every countable subspace of X is compact.*

(1_o) *Every open subspace of X is compact.*

(2), (2_c), (2_o) *Every subspace (or countable, or open subspace) of X is sequentially compact.*

(3), (3_c), (3_o) *Every subspace (or countable, or open subspace) of X is countably compact.*

(4) *X has no weakly discrete infinite subspace.*

(5) *Every strictly decreasing sequence of closed subsets of X is finite.*

(6) *X has a sub-base $\mathfrak{B}$ of open sets such that every strictly increasing sequence of finite unions of members of $\mathfrak{B}$ is finite.*

Remark. The equivalence of (1), (1_o) and (5) is known (see [10] and Exposé 1 (by P. Cartier) of the Séminaire C. Chevalley, vol. 1, 1956-8).

Proof. Because countable compactness is implied by compactness or sequential compactness, it is enough to prove the implications $(3_c) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1)$, $(6) \Rightarrow (5)$, $(3_o) \Rightarrow (1)$ and $(4) \Rightarrow (2)$. All are easy; by way of example we prove $(4) \Rightarrow (5)$. If $C_1 \supset C_2 \supset \cdots$ is an infinite strictly decreasing sequence of closed subsets of X, pick $p_n \in C_n - C_{n+1}$ $(n = 1, 2, \cdots)$; the points p_n are all distinct, so the set $P = \{p_n\}$ is infinite. But each p_n has the neighborhood $P \cap (X - C_{n+1})$ in P, and this consists of the n points $p_1, \cdots, p_n$ only. Thus (4) is contradicted.

Consider now the following modified compactness conditions (all weaker than compactness) on a space X. (The list could easily be extended.)

(A) Every open covering of X has a finite subsystem whose union is dense in X.

(B_1) Every locally finite system of open sets in X is finite.

(B_2) Every locally finite system of disjoint open sets is finite.

(B_3) Every countable covering of X has a finite subsystem whose union is dense in X.

(C_1) Every locally finite open covering of X has a finite subcovering.

(C_2) Every countable locally finite open covering of X has a finite subcovering.

(D_1) Every locally finite open covering of X has a finite dense subsystem.

(D_2) Every countably infinite open covering of X has a proper dense subsystem.

(E_1) Every star-finite open covering of X is finite.

(E_2) Every countably infinite open covering of X by sets each of which meets at most two others has a proper dense subsystem.

(F) Every continuous real-valued function on X is bounded.

Property (B_2) is "feeble compactness" [7]; (B_1) has been called "light compactness" [1]; (F) is "pseudocompactness" [2]. It is easily seen that each of these properties implies the next, and that (B_1), (B_2), (B_3) are equivalent (see [1, 6]), and similarly (C_1) and (C_2), (D_1) and (D_2), and (E_1) and (E_2) are equivalent; it can be shown by examples that there are no other implications between them in general.[2] All except (A) are implied by countable compactness, and are equivalent to it for normal T_1 spaces [2], but not in general. Weaker separation axioms suffice for some other equivalences (for instance, regularity makes (B)—(E) equivalent [3, 4, 5]). But the hereditary forms of all these properties are equivalent, irrespective of separation axioms, as the next theorem shows.

THEOREM 2. *The following statements about an arbitrary space X are equivalent:*

(1) (1_c) *Every subspace (or countable subspace) of X has property* (A).

(2) (2_c) *Every subspace (or countable subspace) of X has property* (F).

(3) *X has no infinite discrete subspace.*

Further, if X is T_1, these statements are equivalent to the statements in Theorem 1.

It follows, of course, that any of the properties (B_1)—(E_2) could replace (A) or (F) here.

Proof. As (A) $\Rightarrow$ (F), it is enough to prove (2_c) $\Rightarrow$ (3) $\Rightarrow$ (1) and that if X is T_1 then (3) implies property (4) of Theorem 1. The first and last of these are trivial; to prove the second, suppose Y is a subspace of X which

[2] Even for T_1 spaces. An example having property (D) but not (C) is given in [1, p. 502] (note, however, that the statement on p. 503 lines 6, 7 is incorrect). It can be modified to give an example satisfying (E) but not (D). The example given at the beginning of § 3 below has property (A) without being countably compact. The usual space of countable ordinals is countably compact, and so satisfies (B), but does not satisfy (A). A suitable union of a sequence of spaces, each of which has no non-constant continuous function, will satisfy (F) but not (E). A T_0 space satisfying (C) but not (B) (from which a T_1 example can be derived by standard technique) is the set of all finite non-empty sets F of positive integers, in which each F has the single neighborhood $U(F)$ = family of non-empty subsets of F. For the properties discussed here (and many others), see [3, 4, 5, 6].

does not have property (A), and let $\{U_\lambda\}$ be an open covering of Y. If no finite subsystem of $\{U_\lambda\}$ is dense in Y, pick $z_1 \in Y$, say $z_1 \in U_{\lambda_1}$; pick $z_2 \in Y - \bar{U}_{\lambda_1}$, say $z_2 \in U_{\lambda_2}$, and generally pick $z_n \in Y - (\bar{U}_{\lambda_1} \cup \cdots \cup \bar{U}_{\lambda_{n-1}})$, say $z_n \in U_{\lambda_n}$. Put $V_n = U_{\lambda_n} \cap Y - (\bar{U}_{\lambda_1} \cup \cdots \cup \bar{U}_{\lambda_{n-1}})$, an open set containing z_n. The sets $V_1, V_2, \cdots$, are disjoint, so $Z = \{z_n\}$ is an infinite discrete subset of Y, contradicting (3).

Remark. In Theorem 2, in contrast to Theorem 1, it is not enough to require that every *open* subspace of X has the relevant properties, even if X is T_1. This is shown by the following example. Let X be the union of two disjoint infinite sets Y, Z; a subset of X is to be open if it is $\emptyset$ or contains all but finitely many points of Z. Then X is a T_1 space and every open subspace of X has property (A), but the closed subspace Y of X is discrete and does not even have property (F).

Further, the T_0 axiom (instead of T_1) would not suffice for the equivalence of the statements in Theorems 1 and 2. For let X be the space of positive integers, with $\emptyset$, X and the sets $\{1, 2, \cdots, n\}$ $(n = 1, 2, \cdots)$ as the only open sets. Every subset of X has property (A), but X is not compact.

3. Irreducibility. A space X is *irreducible* if it is not the union of two proper closed subsets; equivalently, every two non-empty open subsets of X intersect. It is known [10] that a hereditarily compact space is always expressible as the union of a finite number of irreducible sets. Here we amplify this property. We say that a space X is *semi-irreducible* if every family of disjoint (non-empty) open subsets of X is finite. Thus every hereditarily compact space is semi-irreducible; but the converse is false, even for T_1 spaces. (Take, for example, X to be an uncountable set in which the closed sets are X and its countable subsets; X is T_1 and irreducible but not even countably compact.) We note the following easily verified properties:

(1) If $A \subset X$, A is irreducible, or semi-irreducible, if and only if $\bar{A}$ has the corresponding property.

(2) If X is irreducible, or semi-irreducible, then so is every open subset of X.

(3) If X is semi-irreducible and non-empty, then X contains a non-empty maximal open irreducible subspace, and also a non-empty maximal irreducible subspace (which must be closed, from (1)).

(4) X is hereditarily irreducible if and only if the open sets of X are linearly ordered by inclusion; if X is T_1, it is hereditarily irreducible if and only if it has at most one point.

THEOREM 3. *The following statements about an arbitrary space X are equivalent:*

(i) *X is semi-irreducible.*

(ii) *There is a finite system of disjoint open irreducible subspaces $U_1, \cdots, U_n$ of X such that $\bigcup \bar{U}_i = X$.*

(iii) *X is the union of a finite number of disjoint irreducible subspaces, each the difference between two closed sets.*

(iv) *X is the union of a finite number of closed irreducible subspaces.*

(v) *X is the union of a finite number of semi-irreducible subspaces.*

(vi) *There is an integer N such that X does not contain more than N disjoint non-empty open sets.*

(vii) *X has only finitely many regular open sets.*[3]

Proof. (i) $\Rightarrow$ (ii) By Zorn's lemma there is a maximal system $\mathcal{U}$ of disjoint open irreducible subsets of X; from (i), this system is finite, say $\mathcal{U} = \{U_1, \cdots, U_n\}$. Let $V = X - \bigcup \bar{U}_i$; from (2) and (3) above, if $V \neq \emptyset$, V contains a non-empty open irreducible subset U_{n+1}, contradicting the maximality of $\mathcal{U}$. Hence $V = \emptyset$ and $X = \bigcup \bar{U}_i$.

(ii) $\Rightarrow$ (iii) Put $Y_i = \bar{U}_i - \bigcup \{\bar{U}_j \mid j < i\}$ $(1 \leqq i \leqq n)$; then $U_i \subset Y_i \subset \bar{U}_i$, so $\bar{Y}_i = \bar{U}_i$ and Y_i is irreducible by (1) above. Since $X = \bigcup Y_i$, (iii) follows.

(iii) $\Rightarrow$ (iv) If $X = \bigcup Y_i$, where Y_i is irreducible, then $X = \bigcup \bar{Y}_i$, where $\bar{Y}_i$ is irreducible.

(iv) $\Rightarrow$ (v) trivially, because every irreducible space is semi-irreducible.

(v) $\Rightarrow$ (vi) Say $X = X_1 \cup \cdots \cup X_n$, where each X_i is semi-irreducible. Because (i) implies (iii), each X_i is the union of a finite number of irreducible sets, so we may write $X = Y_1 \cup \cdots \cup Y_N$, where each Y_i is irreducible. Suppose that $U_1, \cdots, U_{N+1}$ are disjoint non-empty open subsets of X. Each U_i meets some $Y_{j(i)}$, and we must have $j(i_1) = j(i_2)$ for two distinct integers i_1, i_2 (between 1 and $N+1$). Thus we may assume that both U_1 and U_2 meet Y_1; but this contradicts the irreducibility of Y

The implication (vi) $\Rightarrow$ (i) is trivial.

(ii) $\Rightarrow$ (vii) Let V be any regular open set in X; we show V is one of the 2^n interiors of unions of the sets $\bar{U}_i$ in (ii). We may suppose V meets $U_1, \cdots, U_r$ and is disjoint from $U_{r+1}, \cdots, U_n$ (where $0 \leqq r \leqq n$). Then, if $i \leqq r$, the closure $\mathrm{Cl}(V \cap U_i)$ of $V \cap U_i$ in X must be $\bar{U}_i$; for, as U_i is irreducible, the non-empty open set $V \cap U_i$ is dense in U_i. Hence

[3] A set G is "regular open" if and only if $G = \mathrm{Int}(\bar{G})$.

$$V = \operatorname{Int}(V) = \operatorname{Int}(\bigcup \operatorname{Cl}(V \cap U_i)) = \operatorname{Int}(\bar{U}_1 \cup \cdots \cup \bar{U}_r).$$

(vii) $\Rightarrow$ (i) If X has an infinite family of disjoint (non-empty) open sets $G_1, G_2, \cdots$, the sets $X - \bar{G}_1, X - \bar{G}_2, \cdots$ provide infinitely many distinct regular open sets.

COROLLARY 1. *If X is semi-irreducible (a fortiori if X is hereditarily compact), X has only a finite number of components.*

For an irreducible space is connected.

COROLLARY 2. *If X is regular, X is semi-irreducible if and only if X has only finitely many open sets.*[4]

For, in a regular space, every open set is a union of regular open sets.

Remarks. (a) From (iv) of Theorem 3, we can write any semi-irreducible space X as $X_1 \cup \cdots \cup X_n$, where each X_i is closed and irreducible, and where no X_i is contained in any other. It is easy to see that the sets $X_1, \cdots, X_n$ are then uniquely determined, apart from their order. (Cf. [10] for the hereditarily compact case.)

(b) A connected semi-irreducible T_1 space need not be irreducible.

THEOREM 4. *For any Hausdorff space X, the following assertions are equivalent:*

(I) *X is hereditarily compact.*
(II) *X is semi-irreducible.*
(III) *X is finite.*

For trivially (I) implies (II) and (III) implies (I); that (II) implies (III) follows from Theorem 3(iv) since an irreducible Hausdorff space can have at most one point.

THEOREM 5. *The following statements about an arbitrary space X are equivalent to those in Theorem 2, and thus to the hereditary compactness of X if X is T_1:*

(1) *Every subspace of X is semi-irreducible.*
(1_c) *Every countable subspace of X is semi-irreducible.*

For if every subspace of X has property (A) (§ 2), it is clearly semi-

[4] "Regular" means that each point has a basis of closed neighborhoods; the T_1 axiom is not assumed. In the hereditarily compact case, Corollary 2 is due to Nollet [9].

irreducible. Conversely, if every countable subspace of X is semi-irreducible, X can contain no infinite discrete subspace.

Remark. The analogous statement (1_0)—that every *open* subspace of X is semi-irreducible—would not be equivalent to the statements in Theorem 5 in general, being equivalent to the semi-irreducibility of X.

4. The type of a hereditarily compact space. Let X be a hereditarily compact space, fixed for the moment. We assign, to each closed subspace of X, an ordinal number, its "type," as follows. The empty set (exceptionally) has type -1. When all the closed subsets of X of types $< \alpha$ have been dealt with, and if X has other closed subsets, then (by Theorem 1(5)) X has *minimal* closed subsets not of type $< \alpha$; each of these is said to be ***irreducibly of type*** α. The finite unions of sets irreducibly of type $\leqq \alpha$ are said to be of type $\leqq \alpha$, and a (closed) set of type $\leqq \alpha$ which is not of type $< \alpha$ is of type α. Ultimately all closed subsets of X (including X) are assigned types. The sets which are irreducibly of type 0 are precisely the non-empty ***trivial*** closed subsets of X; if X is T_1 they are the 1-point sets.[5] (Further examples will be given later.) The following two properties follow at once from the definitions.

(1) If $Y_1, \cdots, Y_n$ are closed subsets of X of types $\leqq \alpha$, then $Y_1 \cup \cdots \cup Y_n$ is of type $\leqq \alpha$.

(2) If Y is closed in X and is irreducibly of type α, then every closed proper subset Z of Y is of type $< \alpha$.

In the following statements it is to be understood that Y is a closed subset of X—a restriction which will later be removed.

(3) If Y is of type α, then every closed subset Z of Y is of type $\leqq \alpha$.

For $Y = Y_1 \cup \cdots \cup Y_n$, where Y_i is irreducibly of type $\alpha_i \leqq \alpha$. Then $Y_i \cap Z$ is of type $\leqq \alpha_i$, by (2), so Z is of type $\leqq \alpha$, by (1).

(4) *Y is of type α if and only if $Y = F_1 \cup \cdots \cup F_n$, where F_i is closed and irreducibly of type α_i and* $\max(\alpha_1, \cdots, \alpha_n) = \alpha$.

If Y is expressible in this form, Y has type $\leqq \alpha$ by (1); but if Y is of type $< \alpha$, then each $\alpha_i < \alpha$ by (3), and therefore $\max(\alpha_1, \cdots, \alpha_n) < \alpha$, which is impossible. Conversely, if Y is of type α, the definition shows that $Y = F_1 \cup \cdots \cup F_n$, where F_i is closed and irreducibly of type $\leqq \alpha$, say of

[5] A space Y is "trivial" if its only closed subsets are $\emptyset$ and Y.

type α_i. Let $\beta = \max(\alpha_1, \cdots, \alpha_n)$; thus $\beta \leqq \alpha$. But (1) shows that Y has type $\leqq \beta$; hence $\beta = \alpha$.

(5) *Y is irreducibly of type α, if and only if Y is irreducible and of type α.*

If Y is irreducible and of type α, we express Y as in (4) with n as small as possible. Because Y is irreducible, $n = 1$, and then $Y = F_1$, irreducibly of type $\alpha_1 = \alpha$. Conversely, if Y is irreducibly of type α, suppose $Y = Y_1 \cup Y_2$, where Y_1, Y_2 are proper closed subsets of Y; by (2), Y_1 and Y_2 have types $< \alpha$, and (1) gives a contradiction.

(6) *If Y is of type α, and $\beta < \alpha$, then Y has a closed subset Z of type β.*[6]

As Y is a closed subset of X which is not of type $< \beta$, Y contains a minimal closed subset Z (of X) with this property; and Z is irreducibly of type β, by definition.

(7) *The type of a closed subset Y of X does not depend on the containing space X.*

It is enough to show that if Y has type α in X, Y has type α when the containing space is Y. We prove by transfinite induction on β that if a closed subset Z of Y has type β in X, it has type β in Y, and conversely. For $\beta = -1$ this is clear. Assume it true for all $\beta < \gamma$, where $\gamma \leqq \alpha$. If Z is *irreducibly* of type γ in X, then Z is a minimal closed subset of X which is not of type $< \gamma$ in X; in view of the induction hypothesis, it is also a minimal closed subset of Y which is not of type $< \gamma$ in Y, and so it is (irreducibly) of type γ in Y. If Z is of type γ in X but not necessarily irreducible, if follows from (4) and the preceding that Z is of type γ in Y. The converse is established by substantially the same argument.

We can thus speak of the *type* of a hereditarily compact space Y, independent of the containing space X; it is, of course, a topological invariant of Y. It follows from (7) that, in propositions (2)-(6), the hypothesis that Y is closed in X can be omitted; these propositions apply to arbitrary hereditarily compact spaces Y.

(8) *There exist hereditarily compact T_1 spaces of type α, for every ordinal α.*

To construct one such space X, let A denote the section of ordinals $< \alpha$, beginning with -1 (that is, we count -1 as an ordinal), let I denote any

[6] It follows that, if α is finite, Y has a family of non-empty closed proper irreducible subsets, well-ordered under inclusion and of ordinal α. This need not be true when α is infinite.

infinite set, and put $X = A \times I$. The closed sets in X are defined to be those of the form $(B \times I) \cup F$, where B is an arbitrary section of A (or A itself) and F is an arbitrary finite set. It is easily verified (using Theorem 1(5)) that X is a hereditarily compact T_1 space; and a straightforward transfinite induction on α shows that X is irreducibly of type α.

The hereditarily compact T_1 spaces of type 0 are finite unions of 1-point spaces—that is, they are the finite (non-empty) T_1 spaces. The hereditarily compact T_1 spaces X irreducibly of type 1 are those of the following form: X is an infinite set and its closed subsets are just X and its finite subsets. The hereditarily compact spaces of finite type n are those of "dimension n" in the sense: $n+1$ is the greatest length of any strictly decreasing sequence of irreducible closed non-empty subsets. This agrees with the usual dimension for algebraic varieties in the Zariski topology. For $n > 1$, and still more for infinite types, there are surprisingly many of them; we return to this in § 7. In the next section we show how all hereditarily compact T_1 spaces of type α can be "constructed" if we know enough about those of types $< \alpha$.

5. Dual direct systems. Let $\{F_\lambda, f_\lambda{}^\mu\}$ be a direct system of spaces F_λ (the suffixes λ running over a directed set Λ) and maps $f_\lambda{}^\mu : F_\lambda \to F_\mu$ $(\lambda < \mu)$ subject to the usual rule $f_\mu{}^\nu f_\lambda{}^\mu = f_\lambda{}^\nu$ for $\lambda < \mu < \nu$. We assume further that the maps $f_\lambda{}^\mu$ are *closed.* Let S be the limit space; thus a point of S is an equivalence class $\{x\}$ of representatives $x = \{x_\lambda\}$, where $x_\lambda \in F_\lambda$ for $\lambda > \lambda(x)$ and $f_\lambda{}^\mu(x_\lambda) = x_\mu$ for $\mu > \lambda > \lambda(x)$; two representatives $\{x_\lambda\}$ and $\{y_\lambda\}$ are equivalent if and only if $x_\lambda = y_\lambda$ for $\lambda > \lambda(x, y)$. We give S, *not* the usual direct limit topology, but one which (roughly speaking) uses closed sets instead of open sets. Let f_λ be the usual mapping of F_λ in S, defined as follows: given $x_\lambda \in F_\lambda$ write $x_\mu = f_\lambda{}^\mu(x_\lambda)$ for $\mu > \lambda$, and put

$$f_\lambda(x_\lambda) = \{\{x_\mu \mid \mu > \lambda\}\} \in S.$$

The closed sets of S are to be the intersections of sets of the form $f_\lambda(K_\lambda)$, where K_λ is closed in F_λ; that is, the sets $S - f_\lambda(K_\lambda)$ form a basis of open sets. It is easily verified that, if $\lambda < \nu$, $f_\nu f_\lambda{}^\nu = f_\lambda$, and thence that, if $\nu > \lambda, \mu$, $f_\lambda(K_\lambda) \cup f_\mu(K_\mu) = f_\nu(f_\lambda{}^\nu(K_\lambda) \cup f_\mu{}^\nu(K_\mu)) = f_\nu(K_\nu)$, where K_ν is closed in F if K_λ, K_μ are closed in F_λ, F_μ. Hence this does define a topology on S, the coarsest in which each f_λ is closed. (In general, the mappings f_λ will not be continuous, even if each $f_\lambda{}^\mu$ is continuous.) We call S, with this topology, the "dual direct limit space" of the system $\{F_\lambda, f_\lambda{}^\mu\}$. Clearly S is a T_1 space if each F_λ is T_1.

We are particularly concerned with the case in which each $f_\lambda{}^\mu$ is 1-1 and

continuous (and thus a homeomorphism into); we then call $\{F_\lambda, f_\lambda{}^\mu\}$ an *imbedding system.* In this case the closed proper subsets of S are simply the sets $f_\lambda(K_\lambda)$, where K_λ is closed in $F_\lambda(\lambda \in \Lambda)$, and each f_λ is a homeomorphism into.

THEOREM 6. *The dual direct limit space S of an imbedding system $\{F_\lambda, f_\lambda{}^\mu\}$ of hereditarily compact spaces, each of type $< \alpha$, is hereditarily compact and of type $\leqq \alpha$; it is irreducible providing no $f_\lambda{}^\mu$ is onto, and T_1 if each F_λ is. Conversely, every irreducible hereditarily compact T_1 space of type $\alpha > 0$ is homeomorphic to the dual direct limit space of an imbedding system $\{F_\lambda, f_\lambda{}^\mu\}$, where each F_λ is hereditarily compact, T_1, and of type $< \alpha$, and no $f_\lambda{}^\mu$ is onto.*

In a sense, this theorem determines all hereditarily compact T_1 spaces by transfinite induction over the type; for any such space is a finite union of closed irreducible subsets of no greater type (4(4) and 4(5)).

Proof. Suppose each F_λ is hereditarily compact and of type $< \alpha$. If there could be an infinite strictly decreasing sequence of closed proper subsets $f_{\lambda_n}(K_{\lambda_n})$ of S $(n = 1, 2, \cdots)$, where K_{λ_n} is closed in F_{λ_n}, the sets $f_{\lambda_1}{}^{-1}(f_{\lambda_n}(K_{\lambda_n}))$ would form a strictly decreasing sequence of closed subsets of F_{λ_1}, which is impossible (Theorem 1(5)). Hence S is hereditarily compact. Each proper closed subset of S, being homeomorphic to a closed subspace of some F_λ, is of type $< \alpha$ (by 4(3) and 4(7)); hence S is of type $\leqq \alpha$. If S is reducible, it is the union of two sets of the form $f_\lambda(K_\lambda)$, $f_\mu(K_\mu)$. Take $\nu > \lambda, \mu$: it follows that $f_\nu(F_\nu) = S$, and thence (because the mappings are 1-1) that $f_\nu{}^\rho$ is onto for all $\rho > \nu$. The T_1 property is obvious.

(Conversely, if the direct limit space S of an imbedding system $\{F_\lambda, f_\lambda{}^\mu\}$ is hereditarily compact and of type $\leqq \alpha$, or is T_1, then the same is true of each F_λ; for F_λ is homeomorphic to a closed subspace of S.)

If X is irreducible, T_1, and hereditarily compact of type $\alpha > 0$, let $\{F_\lambda\}$ be the family of its closed proper subsets, ordered by (proper) inclusion; as $F_\lambda \cup F_\mu$ is also a closed *proper* subset, the family is directed.[7] Let $f_\lambda{}^\mu$ be the inclusion map ("identity") for $F_\lambda \to F_\mu$. This defines an imbedding system; let S be its dual direct limit space. It is easily verified that S is homeomorphic to X (the T_1 axiom guarantees that the obvious map of S in X is onto), and that the other properties asserted hold good.

Remark. In the first part of Theorem 6, to ensure the hereditary com-

[7] The T_1 axiom is used here to produce a closed proper subset of X properly containing F_λ and F_μ when $\lambda = \mu$.

pactness of the direct limit S of a direct system of hereditarily compact spaces, we have assumed that each $f_\lambda{}^\mu$ is closed, continuous and 1-1. None of these assumptions can be omitted; nor can the usual (instead of the dual) direct limit topology be used.

6. Standard operations and types.

LEMMA 1. *If every proper closed subset Z of a space X is hereditarily compact and of type $< \alpha$, then X is hereditarily compact and of type $\leqq \alpha$.*

That X is hereditarily compact follows from Theorem 1(5); the rest follows from the way in which types were defined.

THEOREM 7. *If Y is any subspace of a hereditarily compact space X of type α, then Y is hereditarily compact and of type $\leqq \alpha$.*

This is proved by transfinite induction on α. We may assume that the theorem is true for all smaller types, and also (since we may replace X by $\bar{Y}$, in view of 4(3)) that $X = \bar{Y}$. Suppose first that Y is irreducible; then X is also irreducible (3(1)). Any proper relatively closed subset of Y is of the form $Y \cap Z$, where Z is a closed proper subset of X; say Z has type β. Then $\beta < \alpha$ because X is irreducible; the hypothesis of induction then gives that the type of $Y \cap Z$ is $\leqq \beta < \alpha$, and by Lemma 1 the type of Y is $\leqq \alpha$. Finally, if Y is not irreducible, we have $Y = Y_1 \cup \cdots \cup Y_n$, where each Y_i is (relatively) closed and irreducible and of type α_i say. By the result just established, $\alpha_i \leqq \alpha$ $(i = 1, \cdots, n)$; thus the type of $Y = \max(\alpha_1, \cdots, \alpha_n) \leqq \alpha$.

THEOREM 8. *The union X of a finite number of hereditarily compact spaces $Y_1, \cdots, Y_n$ is hereditarily compact; and if Y_i is of type α_i and X of type α, then*

$$\max(\alpha_1, \cdots, \alpha_n) \leqq \alpha \leqq 1 + \alpha \leqq (\Sigma)(1 + \alpha_i).$$

Here $(\Sigma)\alpha_i$ denotes the "natural" sum of the ordinals $\alpha_1, \cdots, \alpha_n$; that is, we express each α_i in the form $\omega_0{}^{\xi_1}k_{i1} + \omega_0{}^{\xi_2}k_{i2} + \cdots + \omega_0{}^{\xi_m}k_{im}$, where the ordinals ξ_j satisfy $\xi_1 > \xi_2 > \cdots > \xi_m = 0$, and $k_{i1}, \cdots, k_{im}$ are positive integers or 0, and define

$$(\Sigma)\alpha_i = \omega_0{}^{\xi_1} \sum k_{i1} + \cdots + \omega_0{}^{\xi_m} \sum k_{im}.$$

(See [11, pp. 363, 364].) When $\alpha_1, \cdots, \alpha_n$ are all finite, this coincides with their ordinary sum.

Note that $1 + \alpha = \alpha + 1$ if α is finite, but $1 + \alpha = \alpha$ otherwise.

Proof. That X is hereditarily compact is obvious, and that $\alpha \geqq \max(\alpha_1, \cdots, \alpha_n)$ follows from Theorem 7. To prove the remaining inequality, we use transfinite induction over the ordered n-ples $(\alpha_1, \cdots, \alpha_n)$ of ordinal numbers (each $\leqq$ some large enough α^*), ordered lexicographically (n being fixed); this is a well-ordered family. It is convenient to count -1 as an ordinal here. Thus the induction starts with each $\alpha_i = -1$; each Y_i is empty, so X is empty and of type $\alpha = -1$ as required. Now suppose that the assertion is true for all $(\alpha_1', \cdots, \alpha_n') < (\alpha_1, \cdots, \alpha_n)$. We first assume that each Y_i is *irreducibly* of type α_i. If Z is any proper closed subset of $X = \bigcup Y_i$, then $Y_i \cap Z$ is for at least one i a proper closed subset of Y_i; hence if $Y_i \cap Z$ has type β_i, we have $\beta_i \leqq \alpha_i$ $(1 \leqq i \leqq n)$, and $\beta_j < \alpha_j$ for at least one value of j. Thus $(\beta_1, \cdots, \beta_n) < (\alpha_1, \cdots, \alpha_n)$, and it follows from the induction hypothesis that the type β of Z satisfies

$$1 + \beta \leqq (\Sigma)(1 + \beta_i) < (\Sigma)(1 + \alpha_i).$$

Hence, by Lemma 1, the type α of X satisfies $1 + \alpha \leqq (\Sigma)(1 + \alpha_i)$.

In the general case, let $Y_i = \bigcup \{Y_{ij} \mid j = 1, 2, \cdots, m(i)\}$, where Y_{ij} is relatively closed and irreducible, and for each of the $m(1)m(2) \cdots m(n)$ choices λ of suffixes, put $Z_\lambda = \bigcap_i \bar{Y}_{i\lambda(i)}$. Then $Z_\lambda \cap Y_i \subset Y_{i\lambda(i)}$, so we have $Z_\lambda \subset \bigcup \{Y_{i\lambda(i)} \mid i = 1, 2, \cdots, n\}$. By Theorem 7 and the case already dealt with, the type γ_λ of Z_λ satisfies

$$(1 + \gamma_\lambda) \leqq (\Sigma)(1 + \text{type of } Y_{i\lambda(i)}) \leqq (\Sigma)(1 + \alpha_i).$$

But $X = \bigcup Z_\lambda$, a finite union of *closed* sets; hence the type α of X satisfies $\alpha = \max(\gamma_\lambda)$, and the desired relation $(1 + \alpha) \leqq (\Sigma)(1 + \alpha_i)$ follows.

Remark. The inequalities in (2) are "best possible," even for T_1 spaces, as can be seen by taking X to be the example constructed to prove 4(8).

THEOREM 9. *The product X of a finite number of hereditarily compact spaces $Y_1, \cdots, Y_n$ is hereditarily compact; and if Y_i is of type α_i and no Y_i is empty, then the type of X is $(\Sigma)\alpha_i$ $(1 \leqq i \leqq n)$.*

It will suffice to prove this when $n = 2$, as then the general result follows by induction over n. As in the proof of Theorem 8 we use transfinite induction over the ordered pairs (α_1, α_2) in lexicographic ordering, and may assume the theorem for products of spaces of types β_1 and β_2 whenever $(\beta_1, \beta_2) < (\alpha_1, \alpha_2)$. Again, as in Theorem 8, we can easily reduce the proof to the case in which Y_1 and Y_2 are irreducible. It readily follows that X is irreducible too. Let Z be any proper closed subset of X; then $X - Z$ contains a set of the form

$U_1 \times U_2$, where U_1, U_2 are non-empty open subsets of Y_1, Y_2. Then $Y_1 - U_1$ and $Y_2 - U_2$ are of types (say) β_1 and β_2, where $\beta_1 < \alpha_1$ and $\beta_2 < \alpha_2$. Hence, by the induction hypothesis, $(Y_1 - U_1) \times Y_2$ and $Y_1 \times (Y_2 - U_2)$ are hereditarily compact and of types $\beta_1(+)\alpha_2$, $\alpha_1(+)\beta_2$. As they are closed in X, Theorem 8 and 4(1) show that their union T is hereditarily compact and of type $\max(\beta_1(+)\alpha_2, \alpha_1(+)\beta_2) < \alpha_1(+)\alpha_2$. But $Z \subset T$, so the same is true of Z; and from Lemma 1 it follows that X is hereditarily compact and of type $\leqq \alpha_1(+)\alpha_2$. To obtain the reverse inequality, suppose (say) $\alpha_2 \neq 0$. For every ordinal $\gamma_2 < \alpha_2$, Y_2 contains a closed proper subset of type γ_2; applying the induction hypothesis again shows that X contains a closed proper subset of type $\alpha_1(+)\gamma_2$. Thus the type of X is greater than $\alpha_1(+)\gamma_2$ for every $\gamma_2 < \alpha_2$, and so is $\geqq \alpha_1(+)\alpha_2$.

If $\alpha_1 = \alpha_2 = 0$ (i.e., to start the induction), Y_1 and Y_2 are trivial spaces; consequently X is trivial too, and so is hereditarily compact and of type 0.

Remark. A product of infinitely many non-trivial spaces is *never* hereditarily compact. For it contains a subspace homeomorphic to $\prod Y_n$ $(n = 1, 2, \cdots)$, where Y_n consists of two points a_n, b_n and (b_n) is open in Y. But this contains the infinite discrete subset $(b_1, a_2, \cdots, a_n, \cdots)$, $(a_1, b_2, a_3, \cdots)$, etc.

THEOREM 10. *Let f be a continuous mapping of a hereditarily compact space X of type α. Then $f(X)$ is hereditarily compact and of type less than $\omega_0^{\alpha+1}$.*

Let $Y = f(X)$. Each $Z \subset Y$ is compact, being a continuous image of the compact set $f^{-1}(Z) \subset X$. To prove the remainder of the assertion, suppose first that X is irreducible. We show by transfinite induction over α that Y has type $\leqq \omega_0^{\alpha}$, if $\alpha \geqq 0$. When $\alpha = 0$, X is trivial; hence Y is trivial, so its type $= 0 < \omega_0^{0}$. In general, if Z is any closed proper subset of Y, which we may assume to be non-empty, let $f^{-1}(Z) = S_1 \cup \cdots \cup S_n$, where each S_i is a non-empty closed irreducible subset of X, necessarily proper. Let S_i have type β_i, and put $\beta = \max(\beta_1, \cdots, \beta_n)$; thus $\beta < \alpha$, because X is irreducible. By the hypothesis of induction, the type of $f(S_i)$ is $\leqq \omega_0^{\beta}$; and by Theorem 8 the type of $Z = \bigcup f(S_i)$ is $\leqq \omega_0^{\beta} n < \omega_0^{\beta+1} \leqq \omega_0^{\alpha}$. Hence, by Lemma 1, Y has type $\leqq \omega_0^{\alpha}$.

In the general case, we have $X = X_1 \cup \cdots \cup X_m$, where X_j is irreducible of type α_j, and $\alpha = \max(\alpha_1, \cdots, \alpha_m)$ (4(4) and 4(5)). By Theorem 8 and the foregoing, the type of Y is $\leqq \omega_0^{\alpha} m < \omega_0^{\alpha+1}$.

Remark. The bound for the type of $f(X)$ here is sharp, even for T_0

spaces, and even if f is 1-1. But if α is finite and $f(X)$ is T_1, its type is $< \omega_0^{\alpha}$, which is now "best possible"; for infinite α, Theorem 10 is sharp even for T_1 spaces and 1-1 mappings. However, if f is *closed* and continuous, it is easily seen that the type of $f(X)$ does not exceed the type of X.

7. Countable spaces. The simplest hereditarily compact spaces are those which have at most countably many closed (or open) sets. Concerning these we have:

THEOREM 11. *Let X be a hereditarily compact space. Then:*

(1) *The family of open subsets of X is countable if and only if X has a countable base.*

(2) *If X is T_0 and has a countable base, then X is countable.*

(3) *If X is T_1, X has a countable base if and only if it satisfies the first axiom of countability.*

Proof. (1) Let $B_1, B_2, \cdots$ be a countable base of open sets. We show that the open sets coincide with the *finite* unions of the sets B_i—which evidently form a countable family. In fact, if U is open, U is a union of sets B_i, and being compact is covered by a finite number of them. "Only if" is trivial.

(2) By (1), X has at most $\aleph_0$ distinct closed sets; but the sets $\bar{x}$ $(x \in X)$ are all distinct.

(3) Assuming X is "first countable," we first show that X is countable, using transfinite induction over the type α of X. We may clearly assume that X is irreducible and not empty. Pick $p \in X$ and let $U_1, U_2, \cdots$ be a basis of open neighborhoods of p; thus $\bigcap U_n = (p)$. Put $F_n = X - U_n$; then F_n has type $< \alpha$, so by the induction hypothesis F_n is countable. Hence X is countable.[8] If the points of X are enumerated as $q_1, q_2, \cdots$, and if $V_{n1}, V_{n2}, \cdots$ is a basis of open neighborhoods of q_n, the sets V_{nm} evidently form a countable basis for X.

The converse implication is trivial.

Remark. There are hereditarily compact T_0 spaces which satisfy the first axiom of countability and have arbitrarily large cardinal, and there are hereditarily compact T_1 spaces (of type 1) which are separable but have arbitrarily large cardinal.

[8] This argument applies, more generally, if instead of assuming that X is T_1 and first countable, we assume that each point of X is a G_δ in X.

One might expect that, conversely, a countable T_1 hereditarily compact space has to satisfy the first axiom of countability, at least at one point, especially in view of a theorem of S. Mrówka [8] asserting that a compact T_2 space with fewer than $2^{\aleph_1}$ points must satisfy the first axiom of countability at some point. But this is not the case, as the following example shows.

Example 1. *There exists a countable hereditarily compact T_1 space of type 2, having c closed subsets, and not having a countable base of neighborhoods at any point.*

The example requires the following lemma, which is due to Sierpinski (cf. [11, p. 77]).

LEMMA 2. *Let S be a set with $\aleph_0$ elements. There is a family of c distinct infinite subsets A_x of S, every two of which intersect in at most a finite set.*

We may take S to be the set of rational numbers, and for each real number x take A_x to be a sequence of rational numbers converging to x.[9]

Now topologize S by taking its closed sets to be: S, and all sets of the form $E \cup \bigcup A_{x_i}$ $(i = 1, 2, \cdots, n)$, where n is a non-negative integer and E is finite. S is easily seen to be irreducible, hereditarily compact and T_1. The closed sets of type 0 are the non-empty sets E; the irreducible closed sets of type 1 are the sets A_x; and thus S is of type 2. We may assume that, given $p \in S$, there are uncountably many sets A_x which do not contain p; for the set of points p for which this is not true must be finite, and we simply omit them from S. If $V_1, V_2, \cdots$ is a countable basis of open neighborhoods of p, we have $V_m = S - (E_m \cup \bigcup \{A_x \mid x \in F_m\})$, where E_m, F_m are finite. As $\bigcup F_m$ is countable, there exists a suffix $y \notin F_m$ for which $p \notin A_y$, and $S - A_y$ is a neighborhhod of p. It must contain some V_m and then $A_y \subset E_m \cup \bigcup \{A_x \mid x \in F_m\}$. But each $A_y \cap A_x$ $(x \in F_m)$ is finite, so A_y is finite, giving a contradiction as required.

A *countable* hereditarily compact space X can evidently have at most c closed subsets; its type must therefore have cardinal $\leqq c$ (from 4(6)). Further, as there are at most 2^c ways of selecting the c sets which are to be closed in X, there can be at most 2^c nonhomeomorphic countable hereditarily compact (or, indeed, countable) spaces. We show now that these trivial estimates are in fact "best possible," even for T_1 spaces. That there can be as many as c closed sets has been shown by Example 1.

[9] This simple proof of Lemma 2 is also due to Sierpinski.

Example 2. *There exists, for each ordinal λ of cardinal $\leqq c$, a countable hereditarily compact T_1 space of type λ.*

We use transfinite induction over λ. Using the sets S, A_x of Lemma 2, and noting that the number of ordinals $\beta < \lambda$ is at most c, we assign to each $\beta < \lambda$ one or more spaces A_x and topologise them as (countable) hereditarily compact T_1 spaces of type β. Now define a topology on S by taking the closed sets to be: S, and all sets of the form $\bigcup F_{x_i}$ $(i = 1, 2, \cdots, n)$, where F_{x_i} is closed in A_{x_i}. One easily verifies that this does give a topology on S, in which S is T_1 and hereditarily compact, and that the subspace topology it induces on each A_x coincides with the topology originally assigned to A_x. Hence S is irreducibly of type λ.

THEOREM 12. *There are 2^c nonhomeomorphic countable hereditarily compact T_1 spaces.*

Let Ω denote the smallest ordinal of cardinal c; let P be the set of ordinals less than Ω, Q the set of non-limit ordinals in P, and R any subset of Q. Thus there are 2^c distinct sets R, and for each of them there are c elements in $P - R$. We construct for each R a corresponding space as follows. Again we use Lemma 2. Let $x \leftrightarrow \alpha(x)$ be a 1-1 correspondence between the set of suffixes x and the set $P - R$. From Example 2, we can give each set A_x a hereditarily compact T_1 topology, irreducibly of type $\alpha(x)$. As in Example 2, we can extend these topologies to a hereditarily compact T_1 topology on S. Now the sets A_x will be precisely the maximal proper irreducible closed subsets of S. Hence the topology of S determines the family of types of the sets A_x, and hence determines R. That is, different sets R give nonhomeomorphic spaces S, and the theorem follows.

It would be interesting to know how many nonhomeomorphic countable hereditarily compact T_1 spaces have a given type α. By a slight modification of the above argument one can show that this number is at least $2^{|\alpha|}$ if $\aleph_0 \leqq |\alpha| \leqq c$, where $|\alpha|$ denotes the cardinal of α.

It would also be interesting to have corresponding estimates for hereditarily compact T_1 spaces of larger cardinals. The above methods can of course be extended, but do not suffice to settle the questions in general; the difficulty is that the analogue of Lemma 2 is false for "most" cardinals (see [12]).

THE UNIVERSITY,
MANCHESTER, ENGLAND.

REFERENCES.

[1] R. W. Bagley, E. H. Connell and J. D. McKnight, Jr., "On properties characterizing pseudo-compact spaces," *Proceedings of the American Mathematical Society*, vol. 9 (1958), pp. 500-506.

[2] E. Hewitt, "Rings of real-valued continuous functions I," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 45-99.

[3] K. Iseki, "On weakly compact and countably compact topological spaces," *Proceedings of the Japan Academy*, vol. 33 (1957), pp. 323-324.

[4] ———, "*AU*-covering theorem and compactness," *ibid.*, pp. 363-367.

[5] S. Kasahara, "On weakly compact regular spaces II," *ibid.*, pp. 255-259.

[6] ———, "A note on some topological spaces," *ibid.*, pp. 453-454.

[7] S. Mardešić and P. Papić, "Sur les espaces dont toute transformation réelle continue est bornée," *Glasnik Matematičko-Fizički Astronomski*, Serija II, vol. 10 (1955), pp. 225-232.

[8] S. Mrówka, "On the potency of compact spaces and the first axiom of countability," *Bulletin de l'Académie Polonaise des Sciences* Sér. Sci. Math. Astr. Phys. 61 (1958), pp. 7-9.

[9] L. Nollet, "A propos de la locale compacité des espaces topologiques," *Bulletin de la Société Royale des Sciences de Liège*, vol. 28 (1959), pp. 5-13.

[10] J. P. Serre, "Faisceaux algébriques cohérents," *Annals of Mathematics* (2), vol. 61 (1955), pp. 197-278.

[11] W. Sierpinski, *Cardinal and ordinal numbers*, Warsaw, 1958.

[12] A. Tarski, "Sur la décomposition des ensembles en sous-ensembles presques disjoints," *Fundamenta Mathematicae*, vol. 12 (1928), pp. 186-205.

IMBEDDING OF HOLOMORPHICALLY COMPLETE COMPLEX SPACES.*

By Raghavan Narasimhan.

1. Introduction.

(1.1) In a previous paper [6], the author proved that any open Riemann surface has a closed, nonsingular imbedding in the complex number space C^3. It is our object, in this paper, to obtain analogous theorems on the imbedding of higher dimensional holomorphically complete complex spaces. (Our terminology is that of [4], [5]; all complex spaces considered are "β-Räume" in the sense of [5].) We prove three theorems which contain the following statements:

1. Any Stein manifold of dimension n admits a one-one, proper, regular holomorphic map in C^{2n+1}.

2. Any holomorphically complete complex space of dimension n admits a one-one, proper, holomorphic map in C^{2n+1} which is regular at every uniformizable point.

3. If locally the holomorphically complete space X can be realized as an analytic set in an open set in C^N ($N > n$ fixed) with the induced structure from C^N, then there is a one-one proper map ϕ of X in C^{N+n} whose image with the induced structure from C^{N+n} is isomorphic to X by means of ϕ.

In all cases, these imbeddings are dense in the space of all holomorphic mappings into the corresponding number spaces, with the compact convergence topology.

These results have the effect of refining the theorem announced by R. Remmert [7] that any holomorphically complete space has a one-one proper map into C^m for some m. They also improve a sharper version of the theorem of Remmert which Professor H. Grauert demonstrated to the author (in a lecture at the Tata Institute of Fundamental Research, Bombay), as he said, by the method of Remmert. Our proofs are based on this version given by Professor Grauert. Since no proof of even the original theorem of Remmert has so far appeared, the proof presented by Professor Grauert is given in § 2

* Received May 10, 1960.

(see Theorem 1). It is an important step in the proof of the main theorems of this paper.

The author takes great pleasure in recording here his indebtedness to Professor H. Cartan for many valuable suggestions and discussions, which led to the formulation and proof of the theorems given here. In particular, the proof of Lemma 1 is due to Professor Cartan, and replaces a more complicated proof of the author.

(1.2) *Some preliminary remarks.* The problem that we consider is that of closed imbeddings of complex spaces in number spaces. By this we mean the existence of a one-one proper holomorphic map into $\boldsymbol{C}^N$. We remark that a *proper map* is one for which the inverse image of a compact set is compact, and that a *regular map* of a complex manifold X is one for which the Jacobian matrix has at any $x_0 \in X$ a rank equal to the dimension of X at x_0.

For a closed imbedding of a complex space X to exist, the following conditions must hold.

i) X is holomorph-convex, i. e. for any infinite discrete set on X, there exists a holomorphic function on X which is unbounded on this set.

ii) Given two distinct points of X, there is a holomorphic function on X taking distinct values at these two points.

iii) X is countable at infinity.

If these conditions are satisfied, we say that X is *holomorphically complete* or that it is a *Stein space.* The properties which we require of such spaces are proved in [4]. We remark that, if X is connected, i) implies iii) (see [4]).

Condition i) is equivalent to the following: for any compact set $K \subset X$, the set

$$\hat{K} = \{x \in X \mid |f(x)| \leqq \sup_{y \in K} |f(y)| \text{ for all } f \text{ holomorphic on } X\}$$

is also compact. We shall denote throughout the paper, the space of holomorphic functions on X and the space of holomorphic mappings of X into $\boldsymbol{C}^N$ by $\mathfrak{H}$ and $\mathfrak{H}^N$ respectively. They are provided with the topology of compact convergence. If $f \in \mathfrak{H}^N$, $f = (f_1, \cdots, f_N)$, we shall set $|f| = \max_i |f_i|$.

It is not always possible to find a proper holomorphic map of a Stein space of dimension n into $\boldsymbol{C}^n$. To prove this, let X be an irreducible K-complete space admitting non-constant bounded holomorphic functions (e. g. bounded domains in $\boldsymbol{C}^n$). Suppose that $\phi: X \to \boldsymbol{C}^n$ is a proper holomorphic

map. Then, if $z \in C^n$, $\phi^{-1}(z)$ is a compact analytic set and so is a finite set (by the maximum principle). Hence, by [9, Satz 23], $\phi(X)$ is an analytic set of dimension n in C^n so that $\phi(X) = C^n$. Thus $X \to C^n$ is a ramified covering. This implies that any holomorphic function f on X satisfies an equation

$$f^m + a_1(z) f^{m-1} + \cdots + a_m(z) = 0,$$

where $a_1(z), \cdots, a_m(z)$ are holomorphic on C^n (the $a_i(z)$ are the elementary symmetric functions of the values of f at the points lying over z). If f is bounded on X, the $a_i(z)$ are bounded on C^n and so are constant; this means that f itself is constant. This cannot be the case since X has non-constant holomorphic functions on it.

Let $U \subset X$ be an open subset of a Stein space X. We say that U *is X-convex* if the following holds:

Let $K \subset U$ be a compact set. Let

$$K' = \{x \in U \mid |f(x)| \leqq \sup_{y \in K} |f(y)| \text{ for all } f \in \mathfrak{H}\}$$

($\mathfrak{H}$ being the space of holomorphic functions on X). Then K' is compact. This is equivalent to saying that the set $\hat{K} \cap U$ is compact.

It is a well known theorem (usually called the Oka-Weil theorem) that if U is X-convex, then any holomorphic function on U can be approximated, uniformly on compact subsets of U, by functions holomorphic on X. This result implies that an open set U is X-convex if and only if U is a Stein space and any holomorphic function on U can be approximated, uniformly on compact sets of U, by holomorphic functions on X.

We make another remark. If K is a compact set of X, then any connected component of $\hat{K}$ meets K. To see this, let A be a connected component of $\hat{K}$ not meeting K. Let V_1, V_2 be disjoint open sets in X, $V_1 \cup V_2 \supset \hat{K}$, $V_1 \supset A$, $V_2 \supset K$. Let W be an X-convex open set of X containing $\hat{K}$ and contained in $V_1 \cup V_2$. Let f be holomorphic on X, $|f(x) - 1| < \frac{1}{2}$ for $x \in A$, $|f(x)| < \frac{1}{2}$ for $x \in K$. Then clearly $|f(x)| > |f(y)|$ if $x \in A$, $y \in K$ so that $A \not\subset \hat{K}$; contradiction. This implies that if U is an open set such that $\hat{K} \cap U$ is compact, where K is a compact subset of U, then $\hat{K} \subset U$. In particular, if U is X-convex and $K \subset U$, then $\hat{K} \subset U$.

The following result is an easy consequence of the definition of X-convexity:

(1.3) If X_1 and X_2 are Stein spaces, $f \colon X_1 \to X_2$ a holomorphic map, then the inverse image of an X_2-convex open set is X_1-convex.

We shall later use this.

2. Decomposition into admissible systems. In this section we shall prove, following H. Grauert, that any Stein space of dimension n can be split into $2n+1$ systems U^λ of relatively compact disjoint open sets $U^\lambda{}_i$ such that the following approximation theorem holds: if f_i is a holomorphic function on $U^\lambda{}_i$, and K_i is a compact set of $U^\lambda{}_i$, (for each i) then there is a holomorphic function F on X approximating f_i on K_i for all i. The proof may be looked upon as generalizing the following simple construction for open Riemann surfaces X. Let T_1 be any triangulation of X. Then there is a triangulation T_2 of X such that the sides of T_2 meet those of T_1 only in points. Then $U^\lambda{}_i$ $(\lambda=1,2)$ are the open triangles of T_λ, $U^3{}_i$ are small parametric discs about the intersections of the sides of T_1 with those of T_2.

We begin with a definition.

A *locally finite* system $\{U_i\}$ of *relatively compact* open sets of X is said to be an *admissible system* if

i) $U=\bigcup_i U_i$ is X-convex, $U_i\cap U_j=\emptyset$ if $i\neq j$,

ii) there exists a sequence of open sets $\{B_\nu\}$, $B_\nu\subset\subset B_{\nu+1}$, $\bigcup_\nu B_\nu=X$ such that for each ν, $B_\nu\cup U$ is X-convex.

We introduce certain conventions which we follow throughout. We remark that the sequence B_ν in ii) may be assumed to have the following property also: if $U_i\cap B_\nu\neq\emptyset$, then $U_i\subset B_\nu$. In fact, we have only to replace B_ν by the union B'_ν of B_ν with the U_i that meet it and pass to a subsequence B'_{ν_k} (to ensure that $B'_{\nu_k}\subset\subset B'_{\nu_{k+1}}$). A sequence satisfying this additional condition, we shall, for convenience, call an *associated sequence* of $\{U_i\}$. It is clear that each B_ν of an associated sequence is X-convex.

The theorem given by Professor Grauert then is as follows:

THEOREM 1. *If X is a Stein space of dimension n, then there exist $2n+1$ admissible systems $\{U^\lambda{}_i\}$, $\lambda=1,\cdots,2n+1$, such that*

$$\bigcup_{\lambda=1}^{2n+1}\bigcup_{i=1}^{\infty} U^\lambda{}_i=X.$$

Proof of Theorem 1. We begin the proof by showing that it is sufficient to prove the existence of an admissible system $\{U_i\}$, such that $A=X-U$, $(U=\bigcup_i U_i)$ is a real analytic set on X which does not contain any point of a given countable set on X. To prove this, we remark that it is easy to deduce from [11, § 8] that the following is true. If M_1 is a real analytic set on X there is a countable set $\{x_m\}$ on M_1 such that if M_2 is real analytic

and does not contain any point x_m, then $\dim(M_1 \cap M_2) < \dim M_1$. It follows from this remark that there are $2n+1$ admissible systems $\{U^\lambda{}_i\}$, $\lambda = 1, \cdots, 2n+1$, so that if $A_\lambda = X - \bigcup_i U^\lambda{}_i$, then $\dim(A_1 \cap \cdots \cap A_k) \leqq 2n - k$ for $k = 1, \cdots, 2n+1$. Clearly then $\bigcap_1^{2n+1} A_\lambda = \emptyset$ and we obtain the theorem.

We proceed to the proof of the existence of an admissible system $\{U_i\}$ with the properties stated. There exists a sequence of open sets $B_\nu \subset\subset B_{\nu+1}$, $\bigcup_\nu B_\nu = X$ such that

$$(*) \qquad B_\nu = \{x \in B_{\nu+1} \mid |\Re\phi_{\nu i}| < 1, |\Im\phi_{\nu i}| < 1, \phi_{\nu i} \in \mathfrak{A}, i = 1, \cdots, k_\nu\}$$

(so that in particular, B_ν is X-convex). To prove this, let K be a compact set with $K = K$, where

$$\hat{K} = \{x \in X \mid |f(x)| \leqq \sup_{y \in K} |f(y)| \text{ for all } f \in \mathfrak{A}\}.$$

Let C be any compact set and suppose that there is an open set U with $K \subset U \subset C$. Since $K = \hat{K}$, if $x \in C - K$, then, there is an $f \in \mathfrak{A}$ with $|f(x)| > 2$, $|f(y)| < 1$ for $y \in K$. From the compactness of $C - U$, it follows that there exist finitely many functions $f_1, \cdots, f_k \in \mathfrak{A}$ such that $\max_i |f_i(x)| > 2$ if $x \in C - U$, < 1 if $x \in K$. If $B = \{x \in C \mid |\Re f_i| < 1, |\Im f_i| < 1\}$, then B is an open set with $\bar{B} \subset U$. The existence of the sequence B_ν satisfying $(*)$ is an easy consequence of this remark.

Let $Q_N(R)$ $(R > 0)$ be the "cube" $\{z \in \boldsymbol{C}^N \mid |\Re z_i| < R, |\Im z_i| < R\}$, $z = (z_1, \cdots, z_N)$. We shall, by convention, suppose that $\boldsymbol{C}^N \subset \boldsymbol{C}^M$ if $M \geqq N$. A "rectangle" in $\boldsymbol{C}^N$ will be a set $\{z \in \boldsymbol{C}^N \mid a_i < \Re z_i < b_i; c_i < \Im z_i < d_i, i = 1, \cdots, N\}$, where $a_i < b_i$, $c_i < d_i$ are real numbers.

Let $m_\nu = k_1 + \cdots + k_\nu$ (see $(*)$). It is clear that the $\phi_{\nu i}$ give a proper map of B_ν into $Q_{k_\nu}(1)$. We remark that if $1 \leqq \alpha_\nu \leqq 1 + \epsilon_\nu$, then the domains $\{x \in B_{\nu+1} \mid |\Re\alpha_\nu\phi_{\nu i}| < 1, |\Im\alpha_\nu\phi_{\nu i}| < 1\}$ satisfy the condition corresponding to $(*)$ if the ϵ_ν are small enough.

Let $\Phi_1 \colon B_1 \to Q_{k_1}(1)$ be the mapping defined by $\phi_{11}, \cdots, \phi_{1k_1}$; let $R_2 > \sup_{x \in B_2} |\Phi_1(x)|$ and let $\Phi_2 \colon B_2 \to Q_{m_2}(R_2)$ be the mapping $(\Phi_1, R_2\phi_{21}, \cdots, R_2\phi_{2k_2})$. Let $R_3 > \sup_{x \in B_3} |\Phi_2(x)|$, $\Phi_3 \colon B_3 \to Q_{m_3}(R_3)$ the mapping $(\Phi_2, R_3\phi_{31}, \cdots, R_3\phi_{3k_3})$, and so on.

Let $Q_\nu = Q_{m_\nu}(R_\nu)$ and let Q'_ν be the cylinder on Q_ν as base in $Q_{\nu+1}$ (in the obvious sense). It is clear that Φ_ν is a proper map of B_ν in Q_ν; moreover, *because of condition* $(*)$, *the inverse image in* $B_{\nu+1}$ *of the rectangle* Q'_ν *by* $\Phi_{\nu+1}$ *is precisely* B_ν.

Let $U_1 = B_1$. The boundary hyperplanes of Q'_1 in Q_2 (i.e. the hyperplanes $\Re z_i = \pm 1$, $\Im z_i = \pm 1$, $i = 1, \cdots, k_1$) split Q_2 into a finite number of open rectangles. Let $U_1, U_2, \cdots, U_{p_2}$ be their inverse images by Φ_2 in B_2. (We remark that U_1 is an inverse image, being that of Q'_1.) Consider next the rectangle $Q_{m_2}(R_3)$. The boundary hyperplanes of $Q_2 = Q_{m_2}(R_2)$ and the hyperplanes used to split Q_2 above decompose $Q_{m_2}(R_3)$ into open rectangles, the decomposition being the same as the original one on Q_2. The cylinders on these rectangles as bases in $Q_3 = Q_{m_3}(R_3)$ we denote by $\rho, \rho', \cdots$. Again, it follows from condition (*) that the inverse images in B_3 by Φ_3 of those of these rectangles $\rho, \rho', \cdots$ whose bases lie in Q_2 give the U_k, $k \leqq p_2$. Let U_k $(p_2 < k \leqq p_3)$ be the inverse images of those rectangles $\rho, \rho', \cdots$ which do not meet Q'_2. It is clear that for $k > p_2$, $U_k \cap B_2 = \emptyset$. We continue this process. Obviously the system $\{U_k\}$ of open sets that one obtains in this way is locally finite and disjoint. Moreover since for $k \leqq p_\nu$, $U_k \subset B_\nu$, each U_k is relatively compact. Let $A = X - \bigcup_k U_k$. Clearly $A \cap B_\nu$ is the inverse image under Φ_ν of the union of finitely many hyperplanes in Q_ν so that A is a real analytic set. Moreover, it is clear that A is defined by means of equations of the following type: $\Re\phi_{\nu i} = \pm a_{\mu\nu}$ $(\mu \geqq \nu)$ $\Im\phi_{\nu i} = \pm b_{\mu\nu}$ (where the $a_{\mu\nu}, b_{\mu\nu}$ depend only on the R_ν). Now, it is clear, e.g. if $\alpha_\nu \leqq 2$, that we may choose R_ν to satisfy the following condition: If

$$\Phi_\nu = (\phi_{11}, \cdots, \phi_{1k_1}, \cdots, R_\nu\phi_{\nu 1}, \cdots, R_\nu\Phi_{\nu k_\nu}),$$

put

$$\Psi_\nu = (\alpha_1\phi_{11}, \cdots, \alpha_1\phi_{1k_1}, \cdots, \alpha_\nu R_\nu\phi_{\nu 1}, \cdots, \alpha_\nu R_\nu\Phi_{\nu k_\nu});$$

then we have $R_{\nu+1} > \sup_{x \in B_{\nu+1}} |\Psi_\nu(x)|$. Thus, we may use $\alpha_\nu\phi_{\nu i}$ instead of $\phi_{\nu i}$ to define the U_k. In this case, the set A is obviously defined by $\Re\phi_{\nu i} = \pm a_{\mu\nu}/\alpha_\nu$, $\Im\phi_{\nu i} = \pm b_{\mu\nu}/\alpha_\nu$. Clearly, the α_ν, $1 \leqq \alpha_\nu \leqq 1 + \epsilon_\nu$ can be so chosen that none of these countably many equations is satisfied at any one of a countable set of points.

To complete the proof of Theorem 1, it remains therefore only to show that $\{U_k\}$ is an admissible system. We show this by proving the following: if Q is a cube in $\mathbf{C}^N$ and G is the complement in Q of finitely many hyperplanes parallel to the coordinate hyperplanes, then G is Q-convex. This is enough; for if this is proved, then one easily sees by means of remark (1.3) at the end of the introduction, that $B_\nu \cup \bigcup_{p_\nu < k \leqq p_{\nu+1}} U_k$ is $B_{\nu+1}$-convex and so, since $B_{\nu+1}$ is X-convex, it is itself X-convex.

Each open rectangle in Q is of course Q-convex. But the union of disjoint Runge domains is not necessarily itself convex, even in the case of open

Riemann surfaces (see [6]). Professor Grauert proved that G is Q-convex by showing that it has a semicontinuous extension to Q. We give, in Lemma 1, a general criterion for the union of disjoint convex sets to be convex. The result required to complete the proof of Theorem 1 is an immediate consequence.

LEMMA 1. *Let X be a Stein space and let X_1, X_2 be disjoint X-convex open sets. Then the following three conditions are equivalent:*

(i) *$X_1 \cup X_2$ is X-convex;*

(ii) *if $K_1 \subset X_1$ and $K_2 \subset X_2$ are compact, there exists a holomorphic function f on X such that $\Re f > 0$, on K_1, $\Re f < 0$ on K_2;*

(iii) *the function which is $+1$ in X_1 and -1 in X_2 can be approximated, uniformly on compact sets of $X_1 \cup X_2$, by holomorphic functions on X.*

Proof. Clearly (i) implies (iii); (iii) implies (i). First we show that (ii) implies (iii). Let $K_i \subset X_i$ be compact and f be as in (ii). Let $C_i = f(K_i)$ in complex z-plane. Then $\Re z > 0$ on C_1 and $\Re z < 0$ on C_2. Since, by the Behnke-Stein theorem [1], the union of these two half-planes is $\boldsymbol{C}$-convex, there is a holomorphic function g in the plane with $|g(z) - 1| < \epsilon$ for $z \in C_1$, $|g(z) + 1| < \epsilon$ for $z \in C_2$. Let $h = g \circ f$. Then, we have

$$|h(x) - 1| < \epsilon \text{ for } x \in K_1, \ |h(x) + 1| < \epsilon \text{ for } x \in K_2$$

which proves (iii).

Finally, (iii) implies (i). Let K be a compact subset of $X_1 \cup X_2$. We have to prove that $\hat{K} \cap (X_1 \cup X_2)$ is compact. Let $K_i = K \cap X_i$; K_i is compact. Moreover, since X_i is X-convex $K'_i = \hat{K}_i \cap X_i$ is compact. We shall show that $\hat{K} \cap (X_1 \cup X_2) = K'_1 \cup K'_2$. Let $x_0 \in X_1 \cup X_2$, $x_0 \notin K'_1 \cup K'_2$. Let $x_0 \in X_1$; then, since $x_0 \in X_1$, $\notin K'_1$, there is a holomorphic function g on X with $|g(x_0)| > \sup_{x \in K_1} |g(x)|$. Let h be holomorphic on X and satisfy

$$|h(x) - 1| < \epsilon \text{ on } K_1 \cup \{x_0\}, \ |h(x)| < \epsilon \text{ on } K_2$$

(which exists by (iii)). Then, if ϵ is small enough, and we put $f(x) = g(x)h(x)$, we clearly have

$$|f(x_0)| > \sup_{x \in K_1 \cup K_2} |f(x)|;$$

this proves (i) and the lemma is proved.

The proof of Theorem 1 is therefore complete.

The following theorem of approximation is an easy consequence of the definition of admissible systems.

THEOREM 2. *Let $\{U_i\}$ be an admissible system on the Stein space X. Let K_i be a compact set of U_i, f_i a holomorphic function in U_i and $\epsilon_i > 0$. Then there is a holomorphic function f on X such that*

$$|f(x) - f_i(x)| < \epsilon_i \text{ for } x \in K_i, \quad i = 1, 2, \cdots.$$

Proof. Let $\{B_\nu\}$ be an associated sequence. Let $U_i \subset B_1$, $i = 1, \cdots, i_1$, $U_i \subset B_{\nu+1} - B_\nu$ for $i_\nu < i \leqq i_{\nu+1}$ $(\nu \geqq 1)$. Let K'_ν be a compact set of B_ν containing $B_{\nu-1}$ and the K_i for $i \leqq i_\nu$. Let $\delta_\nu > 0$, $\sum_{\mu=\nu+1}^{\infty} \delta_\mu < \frac{1}{2} \min_{i \leqq i_\nu} \epsilon_i$. Since U is X-convex, there is an $F_1 \in \mathfrak{H}$ with

$$|F_1(x) - f_i(x)| < \tfrac{1}{2}\epsilon_i \text{ on } K_i (i \leqq i_1).$$

Since $B_1 \cup U$ is X-convex, there is F_2 holomorphic on X with $|F_2 - F_1| < \delta_1$ on K'_1, $|F_2 - f_i| < \frac{1}{2}\epsilon_i$ on K_i for $i_1 < i \leqq i_2$ and so on. Since the limit of holomorphic functions, converging uniformly on compact sets, is also holomorphic (see [5]), $f = \lim_{\nu \to \infty} F_\nu$ is holomorphic on X. One sees easily that $|f - f_i| < \epsilon_i$ on K_i.

We remark that if $\{U_i\}$ is a disjoint family of holomorphically convex open sets which satisfy the conclusion of Theorem 2, then $\{U_i\}$ is locally finite. Also, if K_i is a compact set of U_i $(i = 1, 2, \cdots)$ then there is an admissible system $\{V_i\}$ with $K_i \subset V_i \subset U_i$. In this sense, Theorem 2 characterizes admissible systems.

An immediate consequence of Theorems 1 and 2 is the following result:

Any Stein space of dimension n admits a proper holomorphic map into C^{2n+1}.

In fact, let $\{U^\lambda_i\}$, $\lambda = 1, \cdots, 2n+1$, be admissible systems with $\bigcup_{i,\lambda} U^\lambda_i = X$. Since each U^λ_i is relatively compact, there exist compact sets $K^\lambda_i \subset U^\lambda_i$, with $\bigcup_{i,\lambda} K^\lambda_i = X$. By Theorem 2, there is a holomorphic function f_λ on X with $|f_\lambda(x) - i| < 1$ for $x \in K_i$, $i = 1, 2, \cdots$. Since the set of $x \in X$ with $|f_\lambda(x)| \leqq i$ $(\lambda = 1, 2, \cdots, 2n+1)$ is then contained in $\bigcup_{\lambda=1}^{2n+1} \bigcup_{j=1}^{i+1} K^\lambda_j$, the map $(f_1, \cdots, f_{2n+1})$ is proper.

3. A lemma. In this section, we state and prove a lemma which we shall use in the proof of each of the three main theorems.

LEMMA 2. *Let X be a Stein space. Let N be a positive integer. Suppose that to each compact set K of X is given a set $A(K)$ of functions in $\mathfrak{H}^N$. Suppose that $A(K)$ is dense in $\mathfrak{H}^N$ and that if $K \subset K'$, then $A(K') \subset A(K)$. Suppose moreover that the following condition holds: (α) if $K \subset (K')^0$, $f \in A(K)$, then there is an $\epsilon > 0$ such that if $g \in \mathfrak{H}^N$ and $|f-g| < \epsilon$ on K', then we have $g \in A(K)$.*

Let $\{U^\lambda_i\}$, $i = 1, 2, \cdots$; $\lambda = 1, \cdots, N$ be N admissible systems, let K^λ_i be a compact set of $\{U^\lambda_i\}$. Let $\epsilon > 0$ and $\epsilon_i > 0$, $i = 1, 2, \cdots$. Let $g = (g_1, \cdots, g_N) \in \mathfrak{H}^N$ be given and let C be a compact set of X.

Then, there exists $f = (f_1, \cdots, f_N) \in \mathfrak{H}^N$ such that $|f-g| < \epsilon$ on C, $|f_\lambda - g_\lambda| < \epsilon_i$ on K^λ_i; $\lambda = 1, \cdots, N$; $i = 1, 2, \cdots$ which belongs to $A(K)$ for every compact set K of X.

Proof. Let $\{B^\lambda_\nu\}$ be an associated sequence of the admissible system $\{U^\lambda_i\}$. Let $'K^\lambda_\nu$ be a compact subset of B^λ_ν containing $B^\lambda_{\nu-1}$ and the K^λ_i in B^λ_ν. Let C^λ_ν be a compact neighbourhood of $'K^\lambda_\nu$ in B^λ_ν. Let $K'_\nu = \bigcap_{\lambda=1}^{N} {'K^\lambda_\nu}$, $C_\nu = \bigcap_{\lambda=1}^{N} C^\lambda_\nu$. Suppose that $U^\lambda_i \subset B_1$ for $i \leqq i^\lambda_1$, $U_i \subset B^\lambda_{\nu+1} - B^\lambda_\nu$ for $i^\lambda_\nu < i \leqq i^\lambda_{\nu+1}$. Clearly we have $K'_\nu \subset (K'_{\nu+1})^0$, $\bigcup_\nu K'_\nu = X$. Since $A(K') \subset A(K)$ if $K \subset K'$, if $f \in A(K'_\nu)$ for all ν, then $f \in A(K)$ for any compact K. It is clear that we may suppose that $C \subset K'_1$ and that $\epsilon < \epsilon_i$ for $i \leqq \max_\lambda i^\lambda_1$.

$A(K'_1)$ is dense in $\mathfrak{H}^N$; hence there is $f^1 = (f^1_1, \cdots, f^1_N) \in A(K'_1)$ with $|f^1 - g| < \frac{1}{2}\epsilon$ on K'_1. By condition (α), there is a $\delta_1 > 0$ such that if $F \in \mathfrak{H}^N$ satisfies $|F - f^1| < \delta_1$ on C_1, then $F \in A(K'_1)$. Since $B^\lambda_1 \cup U^\lambda (U^\lambda = \bigcup_i U^\lambda_i)$ is X-convex, there are holomorphic functions $'f^2_\lambda$, $\lambda = 1, \cdots, N$ with $|'f^2_\lambda - f^1_\lambda| < \frac{1}{2}\delta_1$ on C^λ_1, $|'f^2_\lambda - g_\lambda| < \frac{1}{2}\epsilon_i$ on K^λ_i for $i^\lambda_1 < i \leqq i^\lambda_2$. Because $A(K'_2)$ is dense in $\mathfrak{H}^N$, there are holomorphic functions f^2_λ, $\lambda = 1, \cdots, N$, on X such that $|f^2_\lambda - f^1_\lambda| < \frac{1}{2}\delta_1$ on C_1, $|f^2_\lambda - g_\lambda| < \frac{1}{2}\epsilon_i$ on K^λ_i, $i_1 < i \leqq i^\lambda_2$, with $f^2 = (f^2_1, \cdots, f^2_N) \in A(K'_2)$.

Let $0 < \delta_2 < \frac{1}{2}\delta_1$ be such that if $F \in \mathfrak{H}^N$ satisfies $|F - f^2| < \delta_2$ on C_2, then $F \in A(K'_2)$. Continuing this process, we find a sequence $f^\nu = (f^\nu_1, \cdots, f^\nu_N) \in A(K'_\nu)$ satisfying the following conditions: $|f^\nu - f^{\nu-1}| < \frac{1}{2}\delta_{\nu-1}$ on $C_{\nu-1}$, $|f_\lambda - g_\lambda| < \frac{1}{2}\epsilon_i$ on K^λ_i, $i^\lambda_{\nu-1} < i \leqq i^\lambda_\nu$. Let $\delta_\nu < \frac{1}{2}\delta_{\nu-1}$ be such that if $F \in \mathfrak{H}^N$ satisfies $|F - f^\nu| < \delta_\nu$ on C_ν, then we have $F \in A(K'_\nu)$. Since $\delta_\nu < \frac{1}{2}\delta_{\nu-1}$, we have $\sum_{\mu=\nu+1}^{\infty} \delta_\mu < \delta_\nu$. Hence if $f = \lim_{\nu\to\infty} f_\nu$, we have $|f - f^\nu| < \delta_\nu$ on C_ν so that $f \in A(K'_\nu)$ for all ν and so $f \in A(K)$ for *any* compact set K of X. Obviously,

if we choose the δ's small enough, we have $|f-g|<\epsilon$ on C, $|f_\lambda-g_\lambda|<\epsilon_i$ on $K^{\lambda}{}_i$. This completes the proof of the lemma.

4. Imbedding of Stein manifolds. We prove in this section our first main theorem. It is contained in the theorem of the next section but is simpler to prove and depends on sofewhat different ideas.

THEOREM 3. *Let X be a Stein manifold of dimension n. Then the set of one-one regular and proper holomorphic mappings of X into $\mathbf{C}^{2n+1}$ is dense in $\mathfrak{H}^{2n+1}$.*

LEMMA 3. *Let X be a Stein manifold of dimension n, K a compact set on X. Then, the set of $f \in \mathfrak{H}^{2n+1}$ which are one-one and regular on K is dense in $\mathfrak{H}^{2n+1}$.*

Proof. Since K is compact, there exists a map $(\phi_1,\cdots,\phi_k): X \to \mathbf{C}^k$ (for some k) which is one-one and regular in a neighbourhood of K. Let $(g_1,\cdots,g_{2n+1})$ be arbitrary holomorphic functions on X. The map (g,ϕ): $X \to \mathbf{C}^{k+2n+1}$ is then one-one and regular in a neighbourhood of K. It follows from a well known argument of H. Whitney [10] that if we put

$$f_i = \sum_{j=1}^{2n+1} a_{ij} g_i + \sum_{j=2n+2}^{m} a_{ij}\phi_{j-2n-1}$$

$$(m = 2n+1+k) \quad i = 1,\cdots,2n+1,$$

then $(f_1,\cdots,f_{2n+1})$ is one-one and regular in a neighbourhood of K for a set of matrices (a_{ij}) dense in the space of all $(2n+1)\times m$ matrices. (If V is a submanifold of $\mathbf{C}^p$, $p > 2n+1$, not necessarily closed, the set of directions so that the projection of $\mathbf{C}^p$ along this direction onto $\mathbf{C}^{p-1}$ is one-one regular on V is dense in the complex projective space $\mathbf{P}^{p-1}$.) The lemma follows easily from this remark.

(Another proof of this lemma can be given on the same lines as the proof of Lemma 6 in Section 6 below.)

LEMMA 4. *If $f \in \mathfrak{H}^k$ is one-one, regular in a neighbourhood of a compact set K of the complex manifold X and if $g \in \mathfrak{H}^k$ satisfies $|f-g|<\epsilon$ on a neighbourhood K' of K, then if ϵ is small enough, g is one-one and regular on K.*

Proof. Let Δ denote the diagonal of the product space $X \times X$. Clearly, if $\epsilon < \epsilon_0$, then g is regular on K. Moreover, if ϵ_0 is small enough, then there is a neighbourhood of any point of K, depending only on ϵ_0, such that g is

one-one in this neighbourhood. This means that there exists a neighbourhood U of $K \times K \cap \Delta$, depending only on ϵ_0, so that if $|f-g| < \epsilon_0$ on K' and $(x,y) \in U - \Delta$, then $g(x) \neq g(y)$. Let $\delta = \inf_{(x,y) \in K \times K - U} |f(x) - f(y)|$. Then $\delta > 0$; if ϵ_0 is small enough, then clearly $|g(x) - g(y)| \geqq \delta/2 > 0$ if $(x,y) \in K \times K - U$. The result follows,

The proof of Theorem 3 is now easy. Let $A(K) \in \mathfrak{H}^{2n+1}$ be the set of $f \in \mathfrak{H}^{2n+1}$ which are regular and one-one on K. Lemmas 3 and 4 imply that $A(K)$ satisfies all the conditions of Lemma 2. Let now $\{U^\lambda{}_i\}$, $\lambda = 1, \cdots, 2n+1$, be admissible systems with

$$X = \bigcup_{\lambda=1}^{2n+1} \bigcup_{i=1}^{\infty} U^\lambda{}_i.$$

Since each $U^\lambda{}_i$ is relatively compact, there exist compact sets $K^\lambda{}_i \subset U^\lambda{}_i$ with $X = \bigcup_{\lambda,i} K^\lambda{}_i$. Let $\{B^\lambda{}_\nu\}$ be an associated sequence of $\{U^\lambda{}_i\}$, let $g = (g_1, \cdots, g_{2n+1}) \in \mathfrak{H}^{2n+1}$ be given, let C be a compact set of X and let $\epsilon > 0$. We may clearly suppose that $C \subset B^\lambda{}_1$ for each λ. Then, by Theorem 2, there are holomorphic functions $h_1, \cdots, h_{2n+1}$ on X with

$$|h_\lambda - g_\lambda| < \epsilon/2 \text{ on } C, \ |h_\lambda| \geqq i+1 \text{ on } K^\lambda{}_i \text{ for } i \geqq i_0(C).$$

By Lemma 2, there is $f = (f_1, \cdots, f_{2n+1}) \in \mathfrak{H}^{2n+1}$ such that f is one-one, regular on *every* compact set, i.e. on X, satisfying $|f - h| < \epsilon/2$ on C, $|f_\lambda - h_\lambda| < 1$ on $K^\lambda{}_i$ for $i \geqq i_0(C)$. The set of points where $|f_\lambda| \leqq i$, $\lambda = 1, \cdots, 2n+1$, is contained in $C \cup \bigcup_{j \leqq \max(i_0(C), i)} \bigcup_{\lambda} K^\lambda{}_j$ and so is compact. Theorem 3 is proved.

THEOREM 4. *On a Stein manifold X of dimension n, there exist $2n+1$ holomorphic functions f_λ such that if D is X-convex and ϕ holomorphic in D, then ϕ is the limit, uniformly on compact sets of D, of polynomials in the f_λ.*

Proof. Let $f = (f_1, \cdots, f_{2n+1}) : X \to \boldsymbol{C}^{2n+1}$ be a one-one, regular and proper holomorphic map. Since ϕ may be approximated by functions on X, we may suppose that $D = X$. The image M of X under f is a nonsingular, closed submanifold of $\boldsymbol{C}^{2n+1}$. Hence the holomorphic function $\phi \circ f^{-1}$ on M can be extended to a holomorphic function Φ on $\boldsymbol{C}^{2n+1}$ (see [2]). If Φ_k is the sum of the terms of degree $\leqq k$ in the Taylor expansion of Φ about the origin, the Φ_k are polynomials for which $\Phi_k(f_1, \cdots, f_{2n+1})$ converges to ϕ as $k \to \infty$.

5. The case of arbitrary Stein spaces. The next theorem concerns

the imbedding of general Stein spaces; it includes Theorem 3. The proof is however somewhat different.

THEOREM 5. *Let X be a Stein space of dimension n. Then, the set of holomorphic mappings $X \to \boldsymbol{C}^{2n+1}$ which are one-one and proper on X and regular at every uniformizable point of X, is dense in $\mathfrak{H}^{2n+1}$.*

Proof. Let $\phi : X \to \boldsymbol{C}^k$ be any holomorphic map. Let $Y = X \times X$ and let $M'(\phi)$ be the set of $(x, y) \in Y$ with $\phi(x) = \phi(y)$. Clearly, $M'(\phi)$ is an analytic set of Y which contains the diagonal Δ of Y. Let $M(\phi)$ be the union of the irreducible components of $M'(\phi)$ which are not contained in Δ. Then $M(\phi)$ is an analytic set of Y.

Let S be the set of non-uniformizable points of X. Then S is an analytic set in X and $X - S$ is a complex manifold. Let $\rho(\phi, x)$ $(x \in X - S)$ denote the rank of the Jacobian matrix of ϕ with respect to a local coordinate system at x. Let $X'(\phi, m)$ $(m \leqq n)$ denote the set of points of $x \in X - S$ with $\rho(\phi, x) \leqq m$. Clearly $X'(\phi, m)$ is an analytic set in $X - S$. Moreover, by a theorem of Remmert [8, Satz 12], the closure $X(\phi, m)$ of $X'(\phi, m)$ in X is an analytic set in X. Clearly none of its irreducible components is contained in S.

We prove first the following result.

(5.1) Suppose that $\{U_i\}$ is an admissible system, $f_k : X \to \boldsymbol{C}^k (k \geqq 0)$ a holomorphic map satisfying the following two conditions:

a_k) $\dim M(f_k) \leqq 2n - k$

b_k) for $0 \leqq m < n$, we have $\dim X(f_k, m) \leqq n - k + m$.

(Here $\dim B \leqq -1$ means that B is empty.)

Let K_i be a compact subset of U_i and let h be a holomorphic function on X. Let C be a given compact set of X and let $\epsilon > 0$, $\epsilon_i > 0$.

Then, there exists a holomorphic function f on X such that

1) $|h(x) - f(x)| < \epsilon$ for $x \in C$; $< \epsilon_i$ for $x \in K_i$; $i = 1, 2, \cdots$,

2) the map $f_{k+1} = (f_k, f)$ of X into $\boldsymbol{C}^{k+1}$ satisfies a_{k+1}) and b_{k+1}).

[We remark that conditions a_0) and b_0) are empty.]

To prove (5.1), we proceed as follows: Let M_q, $q = 1, 2, \cdots$, be the irreducible components of $M(f_k)$. Choose on each M_q a point $(x_q, y_q) \notin \Delta$. Then $\{(x_q, y_q)\}$ is a discrete set of $Y = X \times X$.

Let X_m be the union of the irreducible components of $X(f_k, m)$ which

have dimension $= n - k + m$ $(0 \leqq m < n)$. No irreducible component of X_m is contained in S and consequently $\dim(X_m \cap S) < n - k + m$. Also, by hypothesis $b_k)$, $\dim X(f_k, m-1) < n - k + m$. Hence every irreducible component $X^p{}_m$ of X_m $(p = 1, 2, \cdots)$ contains a point, say $x^p{}_m$, which does not belong to $S \cup X(f_k, m-1)$. Since $x^p{}_m \notin X(f_k, m-1)$, we must have $\rho(f_k, x^p{}_m) = m$. Moreover, the points $\{x^p{}_m\}$, $m = 0, 1, \cdots, n-1$; $p = 1, 2, \cdots$, form a discrete subset of X.

We now make two remarks.

1°) *The set of holomorphic functions on X separating finitely many distinct points of X is dense in $\mathfrak{H}$.* For if $x_1, \cdots, x_p$ are the given points, then clearly there is an $f_i \in \mathfrak{H}$ with $f_i(x_j) = \delta_{ij}$; the function $\sum \lambda_i f_i = f$ has the property that $f(x_i) = \lambda_i$; hence this set is not empty. If now $g \in \mathfrak{H}$ is given, then $g + \lambda f$ fails to separate the x_i only if λ is one of the values $(g(x_i) - g(x_j))/(f(x_j) - f(x_i))$ $(i \neq j)$; since λ can be chosen arbitrarily small and different from these finitely many values, the result is proved.

2°) *Let $x_1, \cdots, x_p$ be finitely many points of $X - S$ and let $f\colon X \to C^k$ be a holomorphic map with $\rho(f, x_i) = \rho_i < n$. Then the set of holomorphic functions $g \in \mathfrak{H}$ such that the map $f' = (f, g)\colon X \to C^{k+1}$ has $\rho(f', x_i) = \rho_i + 1$ for $i = 1, \cdots, p$ is dense in $\mathfrak{H}$.* In the first place, there are holomorphic functions $z_1, \cdots, z_n$ on X forming a coordinate system at each of the points x_i as one sees by taking suitable linear combinations of coordinate systems at the various points consisting of holomorphic functions on X. Again, if $\mu_1, \cdots, \mu_n$ are suitable constants, then the function $\zeta = \mu_1 z_1 + \cdots + \mu_n z_n$ has, together with the given map f, a rank $> \rho$ at x_i. If $g \in \mathfrak{H}$ is arbitrary, then $g + \lambda\zeta$ together with f has rank ρ_i at x_i for some i only if λ is one of finitely many fixed complex numbers. The result follows since λ may be chosen arbitrarily small and $\neq$ these numbers.

Let $x^p{}_m$ and (x_q, y_q) be the points described before 1°) and 2°). We define a subset $A(K)$ of $\mathfrak{H}$ for each compact set $K \subset X$ as follows: $g \in \mathfrak{H}$ belongs to $A(K)$ if the following two conditions are satisfied: (i) if $x^p{}_m \in K$ and f' is the map (f_k, g), then $\rho(f', x^p{}_m) = m + 1$; (ii) if $(x_q, y_q) \in K \times K$, then $g(x_q) \neq g(y_q)$. It is obvious, that if $K \subset (K')^\circ$ and $| g - g_1 | < \epsilon$ on K', then $g_1 \in A(K)$. This and 1°) and 2°) above imply that $A(K)$ *satisfies the conditions of Lemma 2 with $N = 1$.* Hence, by Lemma 2, if $\{U_i\}$ is the given admissible system, there is a $g \in \mathfrak{H}$ with $| g - h | < \epsilon$ on a given compact set C, $| g - h | < \epsilon_i$ on K_i, $i = 1, 2, \cdots$ such that $g \in A(K)$ for any compact set K. This last condition clearly means just that $g(x_q) \neq g(y_q)$ for *all* q and that the map $f_{k+1} = (f_k, g)$ has a Jacobian of rank $m + 1$ at $x^p{}_m$ for each $m = 0$,

$1, \cdots, n-1$ and all p. We assert that the map f_{k+1} satisfies conditions a_{k+1}) and b_{k+1}). In fact, let M be any irreducible components of $M(f_{k+1})$ which is not contained in Δ. Then $M \subset M_q$ for some q (M_q are the irreducible components of $M(f_k)$ not contained in Δ). But $(x_q, y_q) \in M_q$ and clearly $\notin M$; hence $\dim M < \dim M_q \leqq 2n-k$. Hence $\dim M(f_{k+1}) \leqq 2n-(k+1)$ which is a_{k+1}). If A is an irreducible component of $X(f_{k+1}, m)$ then either A is contained in an irreducible component of $X(f_k, m)$ of dimension $< n-k+m$ or A is contained in an irreducible component $X^p{}_m$ of X_m. But then again $x^p{}_m \in X^p{}_m$, $\notin A$ so that $\dim A < n-k+m$. This is b_{k+1}). (5.1) is completely proved.

Theorem 5 is now easily proved. We choose $2n+1$ admissible systems $\{U^\lambda{}_i\}$, $\lambda = 1, \cdots, 2n+1$, and compact sets $K^\lambda{}_i \subset U^\lambda{}_i$ so that $\bigcup_{i,\lambda} K^\lambda{}_i = X$. Let $f = (f_1, \cdots, f_{2n+1}) \in \mathfrak{H}^{2n+1}$ be arbitrary, C a given compact set and $\epsilon > 0$. Then, by Theorem 2, there is a holomorphic function h_λ ($\lambda = 1, \cdots, 2n+1$) with $|h_\lambda - f_\lambda| < \epsilon/2$ on C, $|h_\lambda| \geqq i+1$ on $K^\lambda{}_i$ for $i \geqq i_0(C)$. Because conditions a_0) and b_0) are empty, there is a holomorphic function ϕ_1 on X with $|\phi_1 - h_1| < \epsilon/2$ on C, < 1 on $K^1{}_i$, so that $\dim M(\phi_1) \leqq 2n-1$, $\dim X(\phi_1, m) \leqq n-1+m$ for $0 \leqq m < n$. We the can find successively functions $\phi_2, \cdots, \phi_{2n+1}$ so that if $\Phi_k \colon X \to \boldsymbol{C}^k$ is the map $(\phi_1, \cdots, \phi_k)$, we have $\dim M(\Phi_k) \leqq 2n-k$, $\dim X(\Phi_k, m) \leqq n-k+m$ for $0 \leqq m < n$ with $|\phi_\lambda - h_\lambda| < \epsilon/2$ on C, < 1 on $K^\lambda{}_i$ for $\lambda \leqq k$. Clearly then the map $\Phi_{2n+1} = \Phi$ is proper and $|\Phi - f| < \epsilon$ on C. Since $\dim M(\Phi) \leqq -1$ and $\dim X(\Phi, m) \leqq -n-1+m \leqq -1$ for $0 \leqq m < n$, Φ is one-one on X and regular on $X-S$. This proves Theorem 5.

From the existence of a proper holomorphic map of a Stein space X into number space, one may deduce the following result:

A closed analytic polyhedron in X (compact set Δ defined as the set of x in a fixed neighbourhood of Δ where $|f_j(x)| \leqq 1$, for finitely many $f_j \in \mathfrak{H}$) is the inverse image *in* X of the closed unit polycylinder in some number space $\boldsymbol{C}^N$ by a holomorphic map $X \to \boldsymbol{C}^N$ (and a corresponding result for open polyhedra).

6. Regular imbedding of Stein spaces. Let X be a complex space and $f \colon X \to \boldsymbol{C}^k$ a holomorphic map. Let $x_0 \in X$. Suppose that f is one-one in a neighbourhood of x_0. Then, there is a neighbourhood U of x_0 such that $f \mid U$ is a proper map of U into a spherical ball $B \subset \boldsymbol{C}^k$ about $f(x_0)$. According to a well known theorem of Remmert [9], $f(U)$ is an analytic set M in B. As such, it has an induced analytic structure from B.

A map $f\colon X \to \boldsymbol{C}^k$ is called *regular* at a point $x_0 \in X$ if it is one-one in a neighbourhood of x_0 and if f induces an analytic *isomorphism* of U onto the complex subspace M of B described above. It is said to be regular on a subset of X if it is regular at each point of the subset.

For complex manifolds, the notion of regularity coincides with the statement that the Jacobian matrix of f has rank at x_0 equal to the dimension of X at x_0.

One may define in the same way a regular map of one complex space into another. We then see easily that the composite of regular maps is regular. In particular, if $f\colon X \to \boldsymbol{C}^k$ is regular at x_0 and F is a map of a neighbourhood of $f(x_0)$ into $\boldsymbol{C}^k$ which is regular at $f(x_0)$, then $F \circ f$ is regular at x_0. Again, if $f\colon X \to \boldsymbol{C}^k$ is any holomorphic map and there exists a regular map F of a neighbourhood of $f(x_0)$ into $\boldsymbol{C}^k$ so that $F \circ f$ is regular at x_0, then f itself is regular at x_0.

For a complex space X to admit a regular map into some number space $\boldsymbol{C}^k$, it is clear that there must exist an integer $N > 0$ such that for each point $x_0 \in X$, there is a holomorphic map $f\colon X \to \boldsymbol{C}^N$ which is regular at x_0. If this condition is satisfied, we shall say that X *is of type* N. If n is the dimension of X, then clearly $N \geqq n$; $N = n$ implies that X is a complex manifold.

Not all complex spaces are of type N for some N; there exist even Stein spaces which are not of "finite type." Our next theorem is an analogue of the imbedding theorem for Stein manifolds in the case of those spaces which are locally of bounded type.

THEOREM 6. *Let X be a Stein space of dimension n. Suppose that there exists an integer $N > 0$ such that the space is locally of type N. Then, if $N > n$ the set of one-one, regular, proper holomorphic mappings of X in $\boldsymbol{C}^{N+n}$ is dense in $\mathfrak{H}^{N+n}$.*

Proof. The proof is via two lemmas which are analogous to Lemmas 3 and 4 in the proof of the imbedding of Stein manifolds. For their proof we require the following result which we state before the lemmas. A proof of the theorem is contained in § 8 of the paper [5] by Grauert and Remmert.

Let D be a domain of holomorphy in $\boldsymbol{C}^n$. Let A be an analytic subset of D with the induced analytic structure from D and K a compact set in D. Then, there is a constant $M = M_K$ depending only on K such that if f is a holomorphic function on A with $|f| \leqq 1$ on A, then there is a holomorphic function F in D, $F = f$ on A such that $|F(x)| \leqq M$ for $x \in K$.

LEMMA 5. *Let X be a Stein space and let $K \subset X$ be a compact subset, $K \subset (K')^0$. Let $f \in \mathfrak{H}^k$ be one-one and regular on K. Then there is an $\epsilon > 0$ such that if $|f - g| < \epsilon$ on K', where $g \in \mathfrak{H}^k$, then g is one-one regular on K.*

Proof. From the proof of Lemma 4, it is clear that we have only to prove the following result:

If f is a one-one proper, regular map of an open set U into a ball $B \subset \boldsymbol{C}^k$, if $x_0 \in U$ and if $|f - g| < \epsilon$, then there is a neighbourhood U' of x_0 depending only on ϵ and not on g (for small ϵ) such that g is one-one and regular on U'.

To prove this, let B' be a smaller ball in B concentric with B. If $z_1, \cdots, z_k$ are the coordinate functions in $\boldsymbol{C}^k$, there is an $\epsilon' > 0$ such that if $H_1, \cdots, H_k$ are holomorphic functions on B with $|H_i - z_i| < \epsilon'$ on B', then $(H_1, \cdots, H_k)$ is one-one and regular in a concentric ball B'' which depends only on ϵ'. If $|f - g| < \epsilon$ on U, then $(f - g) \circ f^{-1}$ is a holomorphic map of $f(U)$ into $\boldsymbol{C}^k$. By the theorem stated above, there is an extension u of $g - f$ to B with $|u| < M\epsilon$ on B'. If $G = z + u$ then clearly we have $|G_i - z_i| < M\epsilon$ on B' $(G = (G_1, \cdots, G_k))$, so that, if $M\epsilon < \epsilon'$, G is one-one regular in B''. By the remark on the composite made at the beginning of the section, $g = G \circ f$ is one-one regular in $U' = f^{-1}(B'')$.

LEMMA 6. *Let X be a Stein space which is locally of type $N > n$. Then the set of $f \in \mathfrak{H}^{N+n}$ which are one-one and regular in the neighbourhood of a compact set $K \subset X$ is dense in $\mathfrak{H}^{N+n}$.*

Proof. It is clear from Lemma 5 that X itself is of type N. It is sufficient, because of Lemma 5, to prove that any $x_0 \in K$ has a neighbourhood U such that the set of $f \in \mathfrak{H}^{N+n}$ which are one-one on K and regular in U is dense in $\mathfrak{H}^{N+n}$. Let V be a neighbourhood of x_0 such that there is a map $\phi: X \to \boldsymbol{C}^n$ so that ϕ is a one-one, regular, proper map of V into a ball $B \subset \boldsymbol{C}^N$ about $\phi(x_0)$. Let $A = \phi(U)$. Let $g: X \to \boldsymbol{C}^p$ be an arbitrary holomorphic map. Let $V(g, m)$ $(0 \leqq m < N)$ be defined as follows. Let $g_1 = g \circ \phi^{-1}$ on A. Let $G': B \to \boldsymbol{C}^p$ be a holomorphic map $= g_1$ on A. Then $x \in V$ belongs to $V(g, m)$ if and only if *for any such extension G', the Jacobian of G'* at $\phi(x)$ has rank $\leqq m$. It is clear that $V(g, m)$ is the inverse image in V of the set of points of B where all the G' have rank $\leqq m$. This latter set is defined in B by a family of holomorphic functions *defined everywhere in B*. It is known (see [3, exp. XV and exp. XI]) that such a set is an analytic set. Consequently $V(g, m)$ is an analytic subset of V. Let B' be a smaller concentric ball in B and let $U = \phi^{-1}(B')$. Let $M(g)$ be the set defined in the proof of Theorem 5, viz. the union of the irreducible

components of the set of $(x,y) \in X \times X$ with $g(x) = g(y)$ which are not contained in the diagonal of $X \times X$.

We make the following remark. If $x \in U$ and $g: X \to C^p$ is a holomorphic map so that $g \circ \phi^{-1}$ has an extension G to B of rank $= r < N$ at $\phi(x)$, then the set of $f \in \mathfrak{H}$ so that the map $g' = (g, f)$ has the property that $g' \circ \phi^{-1}$ has an extension G' to B of rank $= r+1$ at $\phi(x)$ is open and dense in $\mathfrak{H}$. To prove this, let H be a holomorphic function on B such that (G, H) has rank $r+1$ at $\phi(x)$. Let h' be a holomorphic function on X with $|h' - H \circ \phi| < \epsilon$ on V. Then $h' \circ \phi^{-1}$ has an extension H' to B with $|H - H'| < M\epsilon$ on B'. Clearly, if ϵ is small enough, (G, H') has rank $r+1$ at $\phi(x)$ so that a function h' exists with the required property. One sees in the same way that the set is open. If $\psi \in \mathfrak{H}$ is arbitrary, Ψ is an extension of $\psi \circ \phi^{-1}$ to B, then if (G, Ψ) has rank r at $\phi(x)$, $(G, \Psi + \lambda H')$ has rank $r+1$ at $\phi(x)$ if $\lambda \neq 0$ which proves our assertion. It now follows, that the following result holds:

Let $g: X \to C^k$ be a holomorphic map satisfying

α_k) the irreducible components of $M(g)$ meeting $K \times K$ have dimension $\leqq 2n - k$.

β_k) The irreducible components of $V(g, m)$ meeting U have dimension $\leqq n - k + m$ for $0 \leqq m < N$.

Then, the set of $h \in \mathfrak{H}$ such that (g, h) satisfies α_{k+1}) and β_{k+1}) is dense in $\mathfrak{H}$.

The proof of this is the same as that of statement (5.1); of course, since U is relatively compact in V, no admissible systems, or Lemma 2, have to be brought in. We omit the details.

Again, since the conditions α_0) and β_0) are empty, it follows that the set of $f \in \mathfrak{H}^{N+n}$ with $\dim(M(f) \cap K \times K) \leqq 2n - n - N (\leqq -1)$ and $\dim(V(f, m) \cap U) \leqq n - (N+n) + m (< 0$ for $0 \leqq m < N)$ is dense in $\mathfrak{H}^{N+n}$. Lemma 6 follows then from the fact that the second condition implies, by the remarks at the beginning of this section, that f is regular.

It may be remarked that the projection method which we used to prove Lemma 3 could also be used to prove Lemma 6.

Theorem 6 follows from Lemmas 2, 5, 6 in the same way that Theorem 3 followed from Lemmas 2, 3, 4.

As in the case of Stein manifolds, Theorem 6 implies that the polynomials in $N + n$ suitable holomorphic functions on a Stein space of dimension n and

type N are dense in $\mathfrak{A}$. If on a Stein space X, the polynomials in finitely many functions of are dense in $\mathfrak{A}$, then, it is shown easily that X must be of type N for some N.

TATA INSTITUTE OF FUNDAMENTAL RESEARCH, BOMBAY.

REFERENCES.

[1] H. Behnke and K. Stein, "Entwicklung analytischer Funktionen auf Riemannschen Flächen," *Mathematische Annalen*, vol. 120 (1948), pp. 430-461.

[2] H. Cartan, "Idéaux et modules des fonctions analytiques de variables complexes," *Bulletin de la Société Mathématique de France*, vol. 78 (1950), pp. 29-64.

[3] ———, *Séminaire E. N. S.*, 1951/52.

[4] H. Grauert, "Charakterisierung der holomorph-vollständigen komplexen Räume," *Mathematische Annalen*, vol. 129 (1955), pp. 233-259.

[5] ——— and R. Remmert, "Komplexe Räume," *ibid.*, vol. 136 (1958), pp. 245-318.

[6] R. Narasimhan, "Imbedding of open Riemann surfaces," *Göttingen Nachrichten*, No. 7 (1960), pp. 159-165.

[7] R. Remmert, "Sur les espaces analytiques holomorphiquement séparables et holomorphiquement convexes," *Comptes Rendus Hebdomaires des Séances de l'Académie des Sciences*, vol. 243 (1956), pp. 118-121.

[8] ———, "Projektionen analytischer Mengen," *Mathematische Annalen*, vol. 130 (1956), pp. 410-441.

[9] ———, "Holomorphe und meromorphe Abbildungen komplexer Räume," *ibid.*, vol. 133 (1957), pp. 328-370.

[10] H. Whitney, "Differentiable manifolds," *Annals of Mathematics*, vol. 37 (1936), pp. 645-680.

[11] ——— and F. Bruhat, "Quelques propriétés fondamentales des ensembles analytiques réels," *Commentarii Mathematici Helvetici*, vol. 33 (1959), pp. 132-160.

DEFORMATIONS OF COMPACT DIFFERENTIABLE TRANSFORMATION GROUPS.*

By Richard S. Palais and Thomas E. Stewart.†

Let G be a Lie group and M a differentiable $(=C^r, r \geqq 1)$ manifold. We recall that a differentiable action of G on M is a differentiable map $\varphi\colon G \times M \to M$ such that $\varphi(g_1, \varphi(g_2, p)) \equiv \varphi(g_1 g_2, p)$, and if e is the identity of G, then $\varphi(e, p) \equiv p$. By a *deformation* of a differentiable action φ of G on M we mean a one-parameter family φ_t $(t \in I = [0,1])$ of differentiable actions of G on M such that $\varphi_0 = \varphi$ and the map $(g, p, t) \to \varphi_t(g, p)$ of $G \times M \times I$ into M is continuous. If the latter map is differentiable, we say that the deformation φ_t is differentiable. Recall that a deformation of M is a one-parameter family ψ_t $(t \in I)$ of diffeomorphisms of M such that ψ_0 is the identity and $(p, t) \to \psi_t(p)$ is a continuous map of $M \times I$ into M. If the latter map is differentiable, then ψ_t is called a differentiable deformation of M. Given a differentiable action φ of G on M and a [differentiable] deformation ψ_t of M we define a [differentiable] deformation φ_t of φ by $\varphi_t(g, p) = \psi_t(\varphi(g, \psi_t^{-1}(p)))$. We call a [differentiable] deformation of φ [*differentiably*] *trivial* if it can be expressed in this form. Recall that two differentiable actions φ_0 and φ_1 of G on M are called equivalent if there exists a diffeomorphism ψ of M such that $\varphi_1(g, p) \equiv \psi(\varphi(g, \psi^{-1}(p)))$. Thus if a deformation φ_t of an action φ is trivial, then φ is equivalent to φ_t for all $t \in I$. Now it is an easy consequence of a theorem proved by one of the authors [1, Theorem 1] that every differentiable action of a Lie group on a Euclidean space with at least one stationary point can be differentiably deformed into a linear action. On the other hand, there is an example of a differentiable (in fact real analytic) action of the circle group on five dimensional Euclidean space which is known not to be equivalent to a linear action. Also, it is a trivial observation that all actions of the line on the torus defined by left translating by a one parameter subgroup are differentiably deformable into each other. However, they clearly are not all equivalent, some being periodic and others not. Thus from the hypothesis that G is compact or that M is

* Received May 23, 1960.

† The authors are National Science Foundation postdoctoral fellows.

compact we cannot conclude that all differentiable deformations of an action of G on M are trivial. However, we will show that if both G and M are compact, then every differentiable deformation of a differentiable action of G on M is in fact differentiably trivial. First, however, we must recall a fairly well-known relation that exists between differentiable deformations of manifolds and differentiable time dependent vector fields. A differentiable time dependent vector field on M is a family $X(t)$ $(t \in I)$ of vector fields on M such that $(p,t) \to X(t)_p$ is a differentiable map of $M \times I$ into the tangent bundle of M (equivalently, in local coordinates the components of $X(t)_p$ are jointly differentiable in t and the coordinates of p). A differentiable path $\gamma: I \to M$ is called an integral curve of $X(t)$ if its tangent at t is $X(t)_{\gamma(t)}$ for all $t \in I$. This is equivalent to the requirement that $t \to (\gamma(t), t)$ be an integral curve of the vector field X^* on $M \times I$ given by $X^*_{(p,t)} = (X(t)_p, D_t)$. If ψ_t is a deformation of M, then we define a differentiable time dependent vector field $X(t)$ on M by $X(t)_{\psi_t(p)} =$ tangent to $s \to \psi_s(p)$ at $s = t$. The uniqueness theorem for ordinary differential equations implies that ψ_t can be recovered from X, for $t \to \psi_t(p)$ is the integral curve of X starting at p. If M is not compact, then not every differentiable time dependent vector field $X(t)$ arises in this way; in fact, there will not in general be an integral curve of $X(t)$ starting at an arbitrary point p of M and defined for the whole unit interval. However, if M is compact, this pathology cannot arise and every differentiable time dependent vector field on M generates a differentiable deformation of M. We now prove

THEOREM. *If φ is a differentiable action of a compact Lie group G on the compact differentiable manifold M, then any differentiable deformation of φ is differentiably trivial.*

Proof. If φ_t is a differentiable deformation of φ, then $\Phi: G \times M \times I \to M \times I$ defined by $\Phi(g,(p,t)) = (\varphi_t(g,p), t)$ is clearly a differentiable action of G on $M \times I$. If we write Φ_g for the diffeomorphism $(p,t) \to \Phi(g,(p,t))$ of $M \times I$ and π for the projection of $M \times I$ on I, then $\pi\Phi_g = \pi$ for all $g \in G$. Let Δ denote the vector field $(0, D)$ on $M \times I$. Since G is compact, we can "average" Δ over G, i.e. form the vector field Δ^* defined by $\Delta^*_{(p,t)} = \int \delta\Phi_g(\Delta_{\Phi_g^{-1}(p,t)})\,d\mu(g)$ where μ is Haar measure on G. Since Φ is differentiable, Δ^* is a differentiable vector field. By invariance of Haar measure Δ^* is invariant, in the sense that $\delta\Phi_g(\Delta^*_{(p,t)}) = \Delta^*_{\Phi_g(p,t)}$. Because $\delta\pi\delta\Phi_g = \delta\pi$ for all $g \in G$ and $\delta\pi(\Delta_{\Phi_g^{-1}(p,t)}) = D_t$, it follows that $\delta\pi(\Delta^*_{(p,t)}) = D_t$, and therefore $\Delta^*_{(p,t)} = (X(t)_p, D_t)$ for some differentiable time dependent vector field $X(t)$ on M. If ψ_t is the corresponding differentiable deformation of M,

then $(\psi_t(p), t)$ is the integral curve of Δ^* starting at $(p, 0)$. Since Δ^* is invariant, it follows that $\Phi_g(\psi_t(p), t)$ is the integral curve of Δ^* starting at $\Phi_g(p, 0) = (\varphi_0(g, p), 0) = (\varphi(g, p), 0)$. But $\Phi_g(\psi_t(p), t) = (\varphi_t(g, \psi_t(p)), t)$ and the integral curve of Δ^* starting at $(\varphi(g, p), 0)$ is $(\psi_t(\varphi(g, p)), t)$, so $\varphi_t(g, \psi_t(p)) = \psi_t(\varphi(g, p))$. Replacing p by $\psi_t^{-1}(p)$ we get $\varphi_t(g, p) = \psi_t(\varphi(g, \psi_t^{-1}(p)))$. q. e. d.

The notion of a deformation of a general group action on a space can be defined, of course, exactly as above replacing all hypothesis of differentiability by continuity, diffeomorphism by homeomorphism, etc. However, the above theorem is false then even under the most stringent demands short of the assumptions of the theorem (indeed "almost all" actions of compact groups on a sphere can be deformed to linear actions). In a sense this is fortunate since we know that the conclusion of the theorem, i. e. two actions that can be deformed into one another are equivalent by conjugation, is too demanding. Saying that two actions φ, φ' are D-equivalent if there is a deformation φ_t such that $\varphi_0 = \varphi$, $\varphi_1 = \varphi'$ provides us with a partition of all action of a compact group G on M which is conceivably more convenient for many of the theorems of transformation groups. Assuming that the group G is compact and the space M on which it acts is a compact, generalized manifold the following problems seem reasonable:

(1) If φ and φ' are D-equivalent actions of G on M, are the orbit spaces of the same homotopy type?

(2) If φ and φ' are D-equivalent, what is the relation of their orbit structures? In particular, are the fixed point sets of the same homotopy type?

It is not even clear in (2) that if φ has fixed points, then the same must be true of φ'.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] T. E. Stewart, "On groups of diffeomorphisms," *Proceedings of the American Mathematical Society*, vol. 11 (1960), pp. 559-563.